

SECI 2143-01

PROBABILITY STATISTICAL & DATA ANALYSIS

21/22 - 2

GROUP 4

PROJECT 2

FINAL REPORT

Submitted to:

DR. Azurah A Samah

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

**TABLE OF CONTENTS**

1.0 INTRODUCTION ……………………………………………………………………… 3

2.0 SCOPE & OBJECTIVE ………………………………………………………………… 4

3.0 METHODOLOGY …………………………………………………………………….. 5

4.0 DATASET USED ……………………………………………………………………… 7

5.0 DESCRIPTIVE STATISTICS ………………………………………………………….. 8

6.0 INFERENTIAL STATISTICS………………………………………………………….. 9

7.0 STATISTICAL DATA ANALYSIS……………………………………………………… 18

8.0 INTERPRETATION OF RESULTS…………………………………………………….. 20

9.0 WORK COORDINATION…………………………………………………………….. 26

10.0 REFLECTIONS………………………………………………………………………. 27

11.0 CONCLUSION & DISCUSSION……………………………………………………… 28

12.0 REFERENCES……………………………………………………………………….. 30

13.0 APPENDICES………………………………………………………………………… 31

**1.0 INTRODUCTION**

The existence of a digital campus has increased the efficiency of university administration while also providing significant convenience to students, faculty, and staff. The digital management system can collect a large amount of data, which is useful in school administration. In terms of student management on a daily basis, if we can learn more about them, we can implement more effective programmes for different students, allowing us to teach students based on their academic ability and raise the school's educational level. The traditional analysis and management of student behaviour is based on their own personal experience and lacks the learner's individualised cognition. At the same time, it cannot provide detailed guidance to students' learning behaviours, nor can it provide personalised learning situations, and promote learning optimization. So, we conducted this analysis to get information about student’s behaviour recently. In this datasets, we use about 19 variables which are daily study time, college mark, stress level, time spent on social media and video games, gender degree certificate of completion, department, height(cm), weight(kg), 10th mark, 12th mark, hobby, study time preference, salary expectation, degree of students’ likeness towards their degree, possibility of choosing their career based on their degree, time of travel, degree of financial status and degree of doing part-time job. In this analysis, we want to apply the use of statistical analysis skill in the dataset, to prove whether there is a relationship between the data. In this report, we will provide information about the analysis of the datasets in various parts. Starting with the description of the chosen datasets, we state our resources of chosen dataset and the sample size that include the number of respondents, number of variables and the transformation of datasets. Next, we will provide the descriptive statistics which will describe the exploratory analysis on each variable and will explain the type of each variable(nominal, ordinal, interval or ratio) with suitable analysis use. Furthermore, we move into inferential statistics which will describe the inferential statistics for 1 sample test, correlation, regression, and Chi Square test of independence. For the next part, the statistical data analysis will be shown by showing how statistical tools which are R are being used.We also provide the descriptive and inferential statistical analysis using R. We also will include the interpretation of our result that involves a lot of understanding of our datasets and problems. Lastly, for a proper report we include the work coordination part, our reflection, conclusion and discussion also references and appendices of this project.

**2.0 SCOPE & OBJECTIVE**

The study was conducted to meet the following objectives:

1. To perform secondary data collection from open source by understanding the nature of the data and setting up the objectives also determine all the variables to measure in the analysis.
2. To organise the data by performing data pre-processing and transforming the datasets. We also will do the modifications by adding and changing the variables to fill the requirement of RStudio used.
3. To apply a suitable statistical method to analyse the datasets. We will use R statistical software and make a summary which includes “univariate” analysis for each variable of interest and “multivariate” analysis between two or more variables.
4. To analyse the data by exploring the patterns or trends of interest of datasets which are factors, relationship between the variables in the datasets. We will apply all the methods to generate all the hypotheses for all testing.
5. To interpret the results by determining the conclusion, significance and implications of the findings.
6. To synthesise the information and document it in a report and disseminate the report in an individual e-portfolio.

**3.0 METHODOLOGY**

This project is conducted by retrieving a secondary dataset from an internet source. The dataset is retrieved to make inferential statistical data analysis out of it. In order for us to produce the inferential statistics, a step-by-step of data analysis must be followed.

* Steps of Data Analysis:

1. Understanding the nature of the problem

In order for us to understand the explicit expectations for the classroom behaviour, an analysis must be conducted. We are directed to the idea of comparing and digging out students’ qualities in excelling in their academics as it is very crucial for one’s bright future. We are looking for information such as whether students have or have not completed their studies, their college marks, total time spent on studying, total time spent on social media and what not.

2. Deciding what to measure and how to measure it

For this research, we will mainly focus on discovering their unequal mean, test statistics, test statistics comparison, the degree of the freedom and relationship between variables. We came up with four different inferential statistical data analysis to prove all the measurements.

3. Data collection

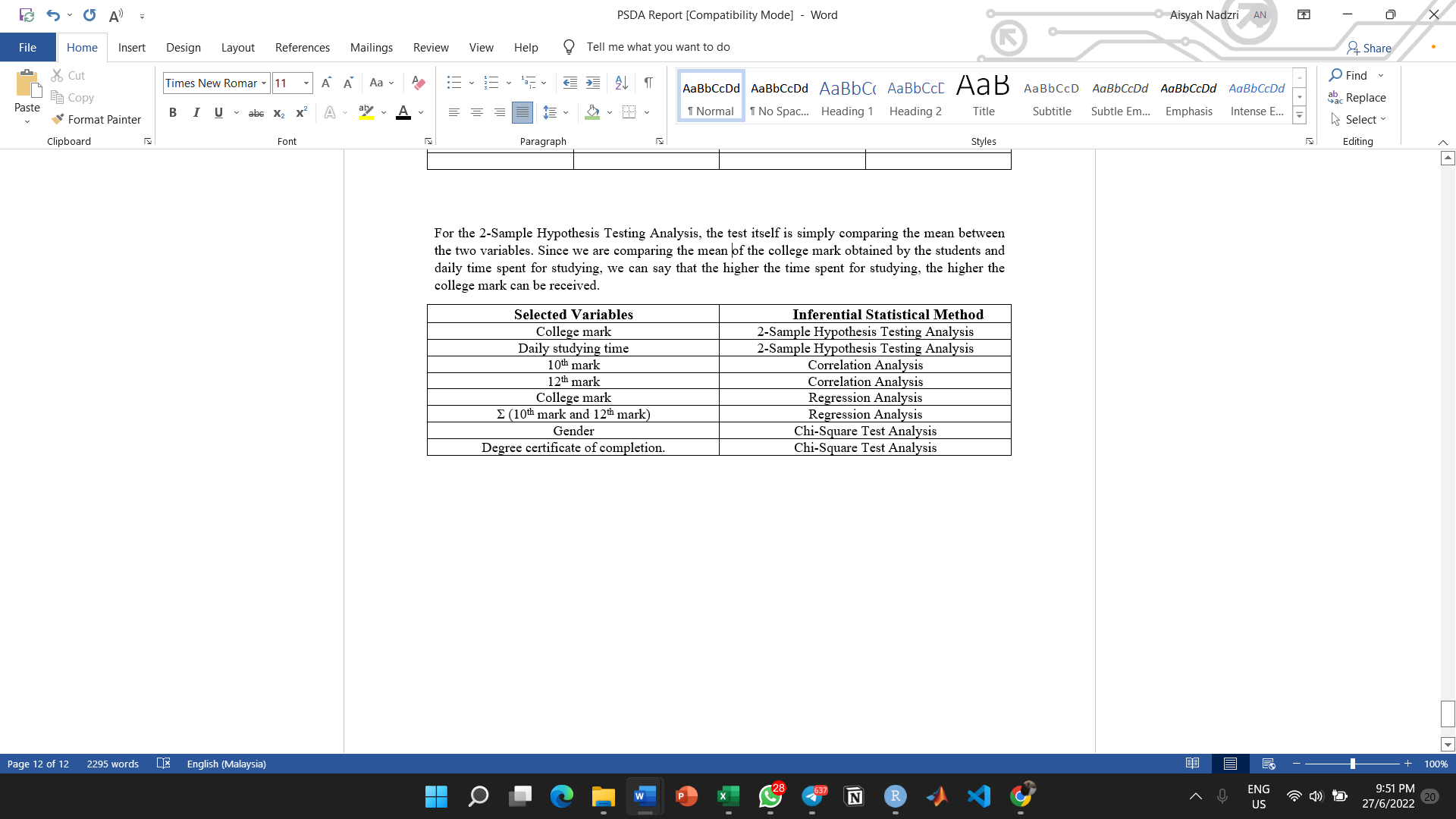
We found a secondary dataset that suits our needs. Before we renamed it, the actual name to the dataset collected was ‘Student\_Behaviour’. We retrieved the dataset from Kaggle.com. It has a size of 235 samples and 19 variables in total. The background of the dataset collected is the dataset contains information from 200 and more university students.

4. Data summarization and preliminary analysis

We expect to perform and produce data summarization numerically and graphically from suitable selected variables. From the dataset collected, we detect a few of high potential variables to make use of.

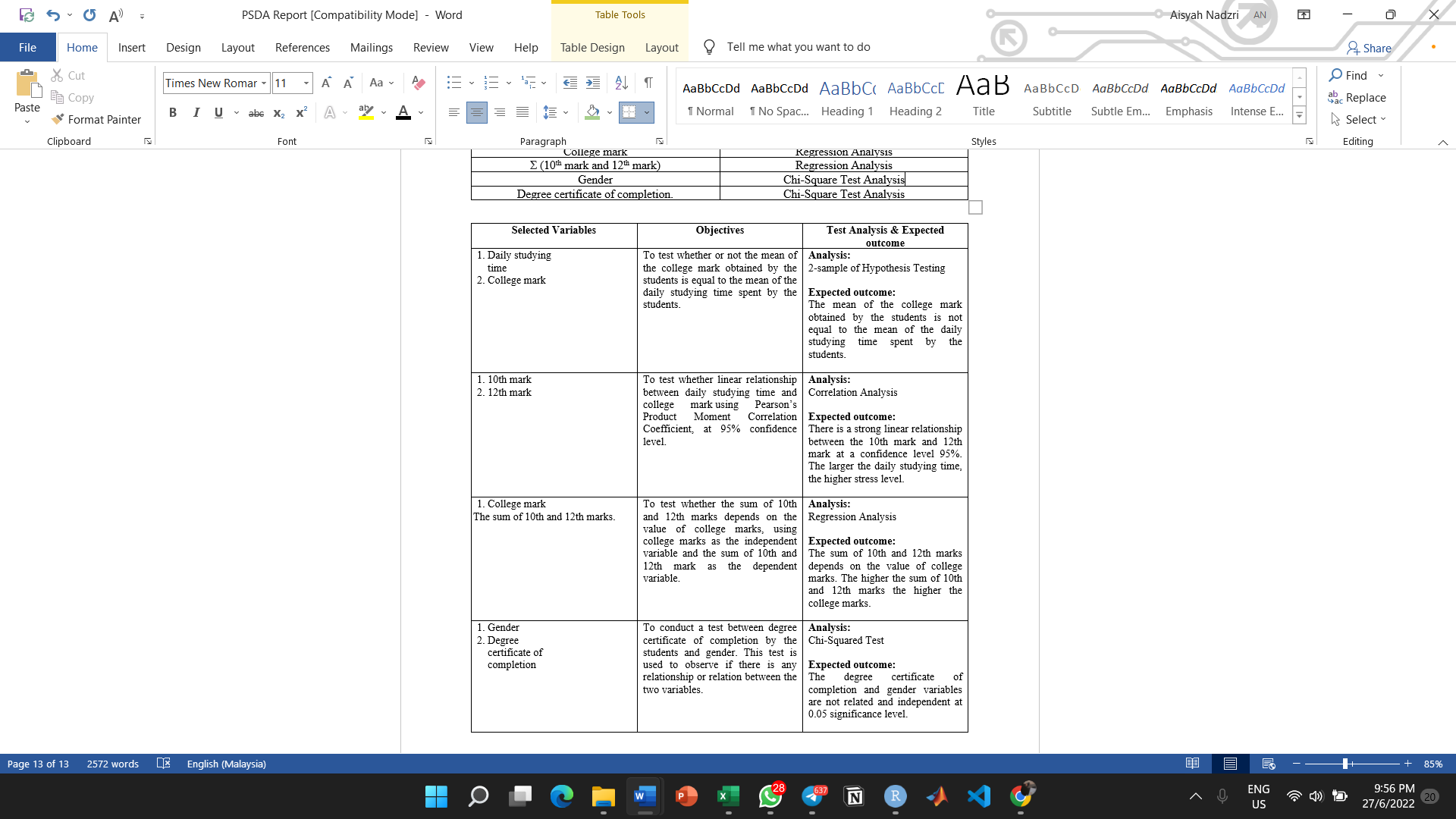
5. Formal Data Analysis

Based on the high potential variables we choose, we can apply the appropriate inferential statistical method. Below is the initial summary of the selected variables and their appropriate inferential statistical method.



6. Interpretation of results

We expect to obtain the results as below



**4.0 DATASET USED**

4.1 Dataset Source

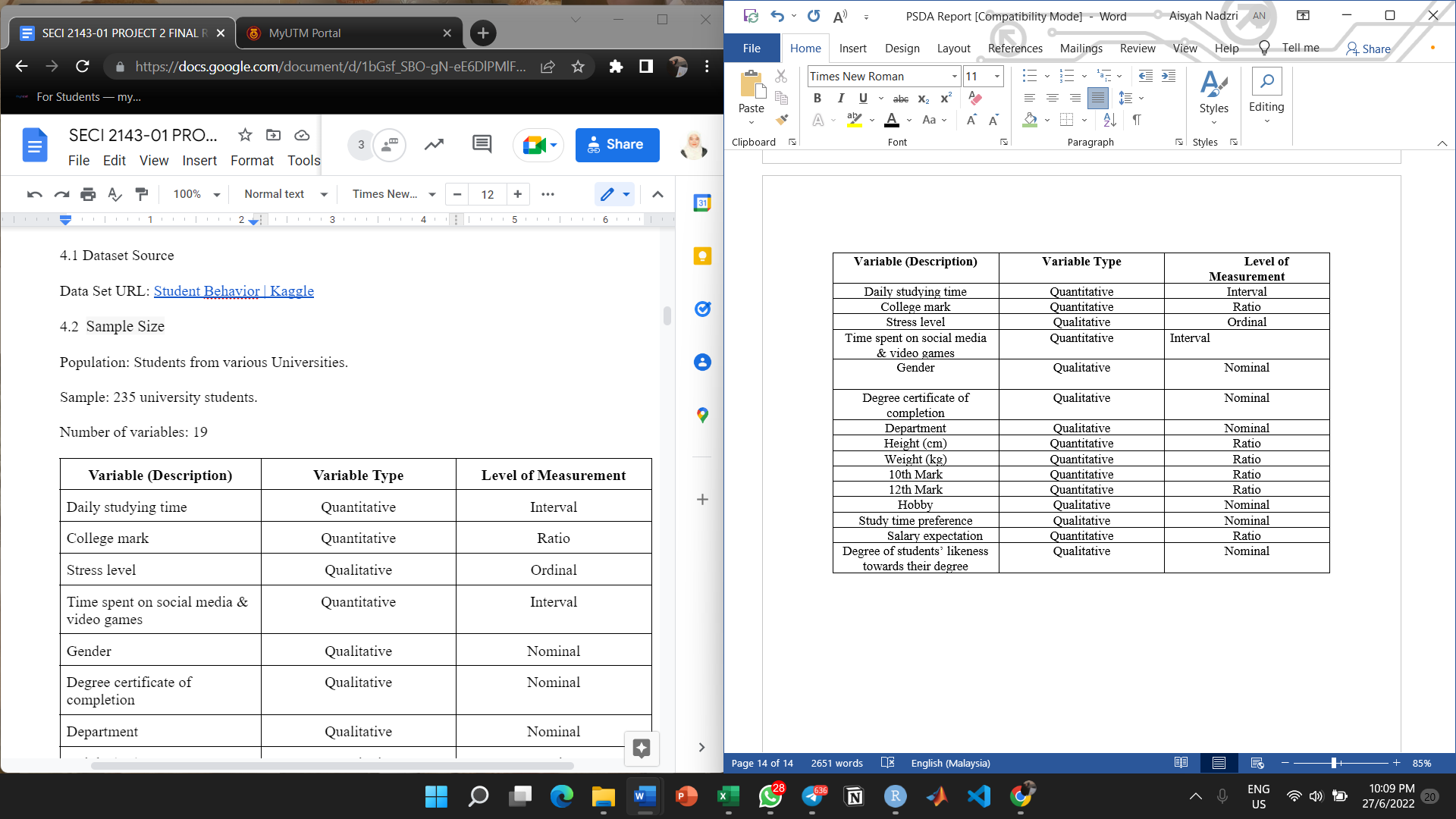
Data Set URL: [Student Behavior | Kaggle](https://www.kaggle.com/datasets/8bde35079db4807b9a5b8e5bfc6c864e5f5b3143544af7a2b67bdd032b1fab59)

4.2 Sample Size

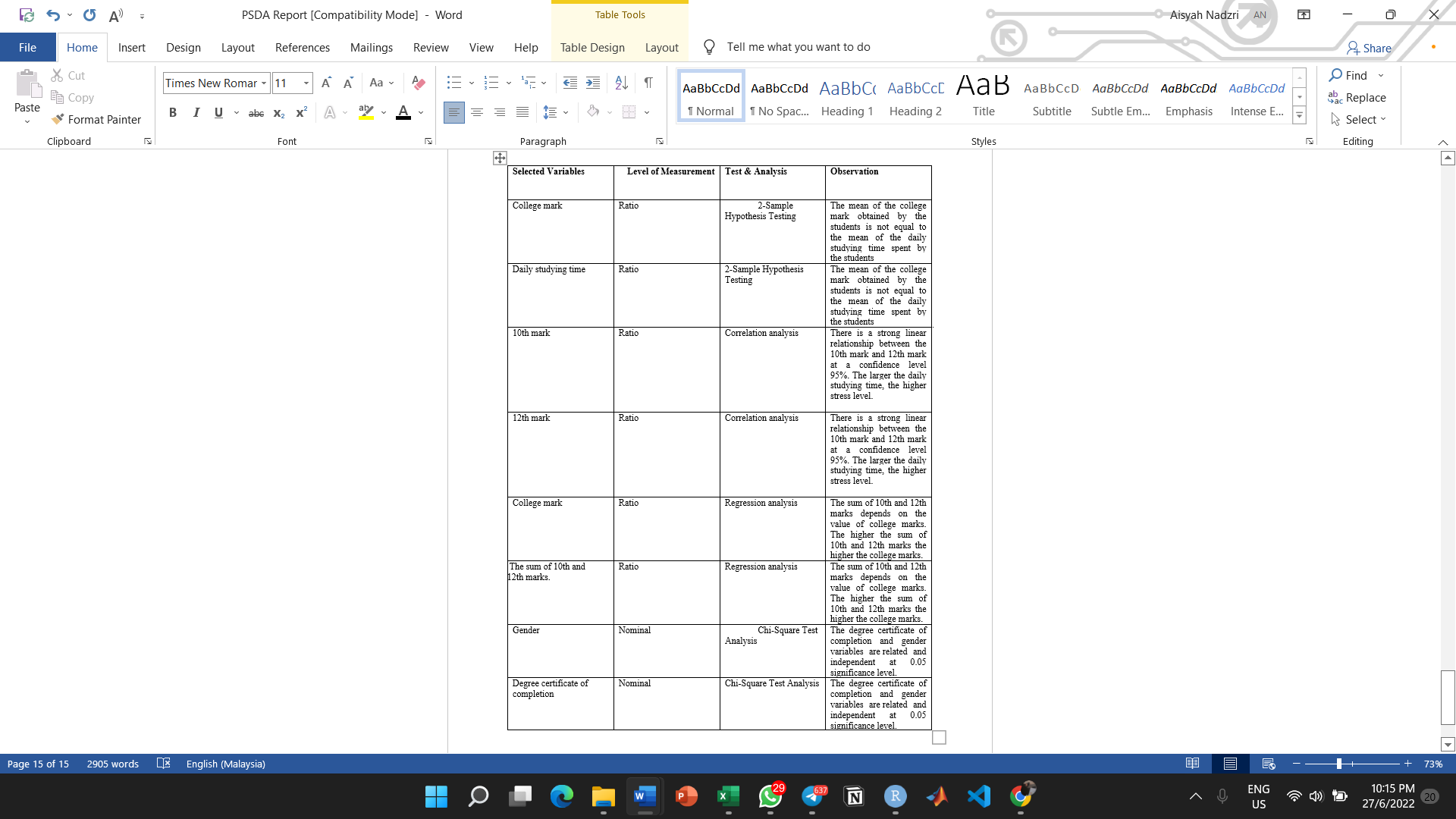
Population: Students from various Universities.

Sample: 235 university students.

Number of variables: 19



**5.0 DESCRIPTIVE STATISTICS**



**6.0 INFERENTIAL STATISTICS**

**6.1 2 - Sample Hypothesis Testing Analysis**

In this research, we wish to determine whether or not the mean of the college mark obtained by the students is equal to the mean of the daily studying time spent by the students, under the t-test 0.05 significance level. The mean of the college mark obtained by the students is 70.66 and 79.02 for the mean of the daily studying time spent by the students while as for the standard deviation, 15.72745 for the college mark obtained by the students and 60.76524 for the daily studying time spent by the students. This is to observe whether the daily studying time spent by the students contributes to the college mark obtained by the students. The variables used in this test are college marks and time spent for studying. Through this observation, an assumption is made where both the mean of these variables are not equal. This test is done by using RStudio.

Let 1 = the mean of the college mark obtained by the students

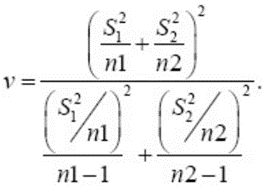
Let 2 = the mean of the daily studying time spent by the students

Hypothesis statement:

H0 : 1 = 2

H1: 1 ≠ 2

In this case we will be using a 95% confidence interval. As for the degree of freedom, it can be calculated by using this formula



and by using RStudio, the degree of freedom, v = 265.

At 95% confidence level, significance level, = 0.05 and the degree of freedom, v = 265, the value of the test statistic can be found in the t-table. The t0.025, 265 is 1.969.

Therefore, using = 0.05 significance level, we reject the null hypothesis if

t\*0 > t0.025, 265 = 1.969

Or

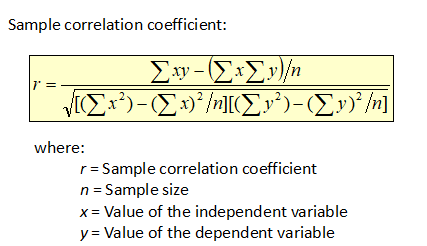
t\*0 < - t0.025, 265 = - 1.969

Since t0 = -2.0419 < t0.025, 265 = -1.969, **we reject the null hypothesis, H0**. There is sufficient evidence to conclude that the mean of the college mark obtained by the students is not equal to the mean of the daily studying time spent by the students.

**6.2 Correlation Analysis**

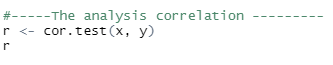
For this analysis, we use two variables from the datasets which are “10th Mark” and “12th Mark”. Correlation analysis is used to measure strength of the association (linear relationship) between two variables. So, in this analysis we want to measure whether the 10th mark and 12th mark has a linear relationship using Pearson’s Product-Moment Correlation Coefficient, at 95% confidence level. We use correlation analysis to test the relationship between these two variables since the 10th mark and 12th mark are ratio-type data.

We can calculate the correlation coefficient using **Pearson’s Product-Moment**. The formulas that we used are shown below.



**Diagram**

We used RStudio to calculate the correlation. You can refer to Diagram 3.



**Diagram**

Now, we want to calculate the significance test for the correlation.

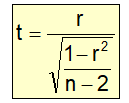
**The hypothesis statement:**

H0: ρ = 0 (no linear correlation)

HA: ρ ≠ 0 (linear correlation exists)

**Test statistic :**

Using formula :



**Diagram 5**



**Diagram**

Inserting formula above in RStudio to get the value of t.

The result is **t = 8.2085 , df = 233, *α* = 0.05**



**Diagram**

From t-table, since this is a two-tailed test, there are two critical values:



**Diagram**

**Lower tail critical value :**

*-tα/2=0.025, df=233* = -1.970198

**Upper tail critical value :**

*tα/2=0.025, df=233* = 1.970198

From RStudio, we also get ***p-value = 1.53e-14***

Hence, if test statistics > 1.970198 or test statistics < -1.970198, reject H0.

Otherwise fail to reject H0.

Since test statistics *t* = 8.2085 > upper tail critical value *tα/2=0.025, df=203* = 1.970198,

we **reject** the null hypothesis. There is **enough evidence** to conclude that there

is a linear relationship between 10th mark and 12th mark, at *α* = 0.05.

**6.3 Regression Analysis**

In the regression test, we want to test whether the sum of 10th and 12th marks depends on the value of college marks, using college marks as the independent variable and the sum of 10th and 12th mark as the dependent variable. We assume the confidence level to be 95%, significant level, α = 0.05.

**Hypothesis statement:**

H0: =**β**0 0 (no linear regression)

H1: **≠β**1 0 (linear regression exists)

Estimated Regression Model: Y = b0+ b1x

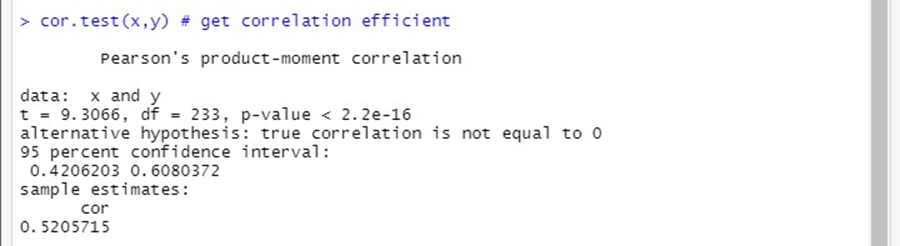
where,

Y = Estimated (or predicted) Y value

b0= Estimate of the regression intercept

b1= Estimate of the regression slope

X = Independent variable

****

**Figure C: Correlation Calculation of College Marks against The Sum of 10th and 12th Marks**

**Test statistical for Regression**

Test statistic, t = ( b1 - **β**1 ) / sb1

where,

b1 = Sample regression slope coefficient slope

sb1 = Estimator of the standard error of the slope

**β**1 = Hypothesised slope

Degree of freedom, d.f = 233

Based on Figure C, we can get the test statistic, t = **9.3066** and P-value = **2.2 x 10-16** From here, we can see that P-value is < than significant level, α = **0.05** . Hence, we reject the null hypothesis.

Hence, there is sufficient evidence that linear regression does exist between the value of college marks and the sum of 10th and 12th marks.

**6.4 Chi - Square Test Analysis**

For this analysis, we are using variables “Gender” and “Degree Certificate of Completion” where we will test whether these two variables are related by using Two Way Contingency Table at α=0.05. Hence, we used RStudio to perform Chi-Square Test of Independence with two-way contingency table.

**Hypothesis Statement:**

H0: = No relationship between Gender and Degree Completion of Certificate

H1: = Gender and Degree Completion of Certificate are related and dependent.

**Critical Value:**

Critical value χ2 = 3.841 (with df=(2-1)(2-1)=1, α = 0.05)

**Expected Counts:**

| **Gender** | **Completed any certificate** | | | | **Total** |
| --- | --- | --- | --- | --- | --- |
| **No** | | **Yes** | |
| **Obs.** | **Exp.** | **Obs.** | **Exp.** |
| **Male** | 64 | (155)(83)/235  = 54.74 | 91 | (155)(152)/235  =100.26 | 155 |
| **Female** | 19 | (80)(83)/235  =28.26 | 61 | (80)(152)/235  =51.74 | 80 |
| **Total** | 83 | 83 | 152 | 152 | 235 |

\*Remarks: eij 5 in all cells.

**Calculated Test Statistics Value:**

Manual Calculation:

| **Cell, ij** | **Observed Count, Oij** | **Expected Count, eij** | **(Oij - eij)^2/eij** |
| --- | --- | --- | --- |
| 1,1 | 64 | (155)(83)/235  = 54.74 | 1.57 |
| 1,2 | 91 | (155)(152)/235  =100.26 | 0.86 |
| 2,1 | 19 | (80)(83)/235  =28.26 | 3.034 |
| 2,2 | 61 | (80)(152)/235  =51.74 | 1.66 |
| **χ2=** | | | 7.124 |

\*When we calculate test statistics manually, we get test statistic χ2 =7.124.

**Decision and Conclusion**

Since the test statistic value (χ2 = 7.124) > critical value ( χ2*k=1, α* = 0.05 = 3.841), it falls within the critical region. Thus, we reject H0 . There is sufficient evidence to conclude that there is a relationship between the variables Gender and Degree Completion of Certificate, at α= 0.05.

**7.0 STATISTICAL DATA ANALYSIS**

7.1 Statistical Tool Used (R)

This project would not have been successful nor completed if it was not for R. R is a language and environment for statistical computing and graphics. It is a [GNU project](http://www.gnu.org/) which is similar to the S language and environment. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible. One of R’s strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. There is a lot to unpack about this specific programming language, because R is an integrated suite of software facilities for data manipulation, calculation and graphical display, It also includes a variety of benefits such as an effective data handling and storage facility, a large, coherent, integrated collection of intermediate tools for data analysis as well as a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

The effectiveness of R in terms of data analysis or data science is incredibly immense. As R provides objects, operators and functions that allow users to explore, model and visualise data. R in data science is used to handle, store and analyse data. It can be used for data analysis and statistical modelling. R is a powerful tool and can be used for bivariate analysis using various inferential statistics. Various other uni-variate and bi-variate analysis can be performed using Descriptive Statistics and that has been explored in 7.2 Descriptive Statistical Analysis using R which can allow us to better understand the data.

7.2 Descriptive Statistical Analysis using R

Not only is R very useful in data analysis related projects, R provides a wide range of functions for obtaining summary statistics. For instance, in the case of **2-Sample Hypothesis Testing Analysis**, one of the many possible functions used includes mean and sd (standard deviation) as well as those unused which comprise of var (variance), min, max median, range and quantile.

Furthermore, in the **Correlation Analysis**, there are numerous R packages that are designed specifically in order to provide a range of descriptive statistics at once. In particular, the stats, psych, ggpubr and rstatix package. Another R function used in the Correlation Analysis is the cor.test(); how it works is that it returns both the correlation coefficient and the significance level(or p-value) of the correlation.

What’s more, in the event of a **Regression Analysis**, some of the implementation of R programming is the usage of Linear Regression which can be translated in R as the function lm(). Aside from the Linear Regression, one implementation of Logistic Regression in R programming is the summary(test) function.

7.3 Inferential Statistical using R

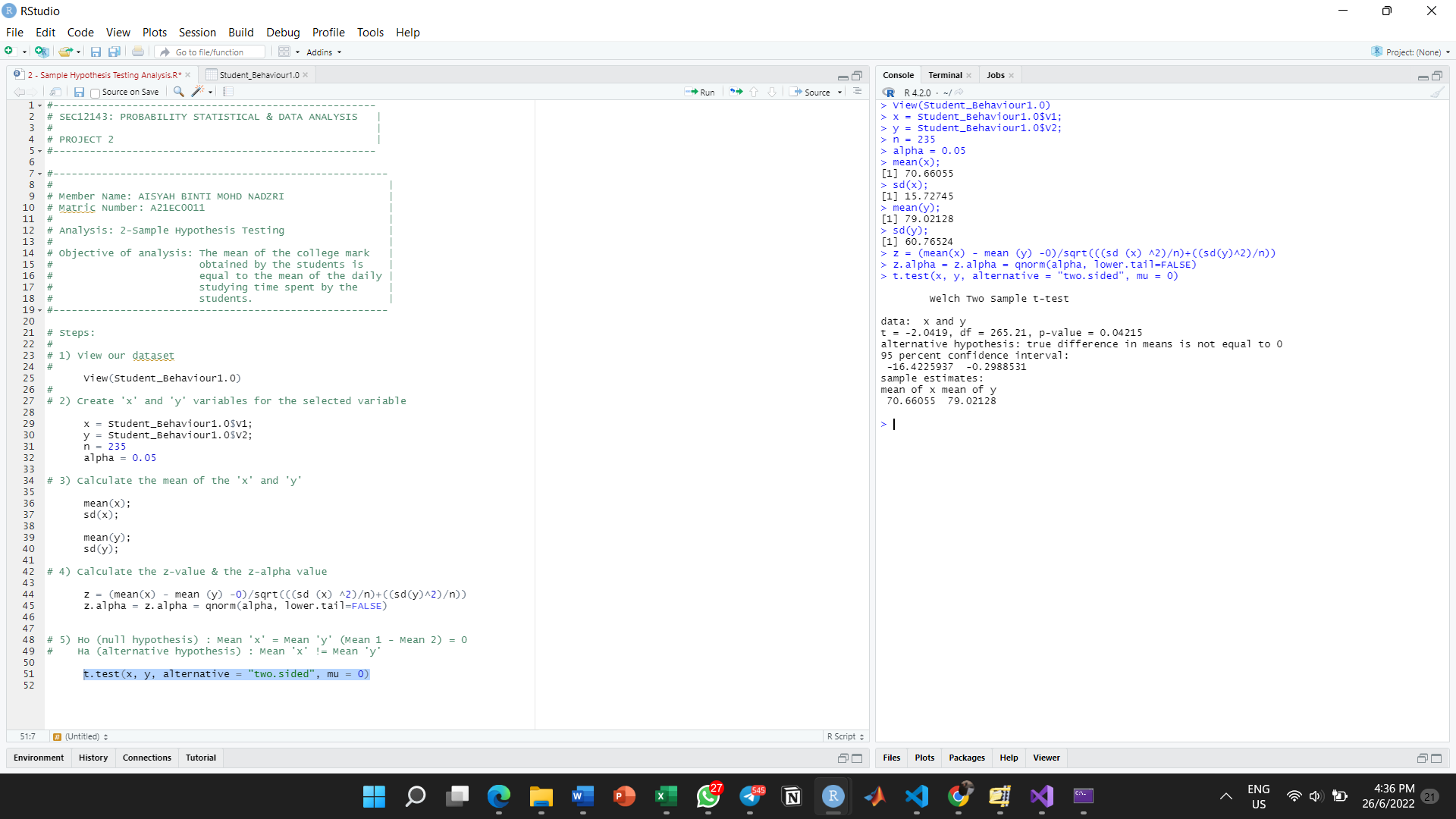
Inferential statistics are used to draw inferences from the sample of a huge data set. Random samples of data are taken from a population, which are then used to describe and make inferences and predictions about the population. One of the most perfect examples for Inferential Statistical using R is **Chi-Square Analysis**. One of the syntax used is chisq.test(data). The distinct syntax functions to perform the chi-square test of independence in the native stats package that are already implemented in R.

**8.0 INTERPRETATION OF RESULTS**

**Analysis:**

2-Sample Hypothesis Testing Analysis

**Result:**



As can be seen, both of the mean for the variables we obtained, which are college marks and daily time spent studying, are different from each other.

Let 1 = the mean of the college mark obtained by the students

Let 2 = the mean of the daily studying time spent by the students

The 1 and 2 are both **70.66** and **79.02** respectively. The difference can be seen perceptibly. In order to get the t-critical value, the degree of the freedom and its proportion value and by following the above R syntax, the calculation steps must be done.

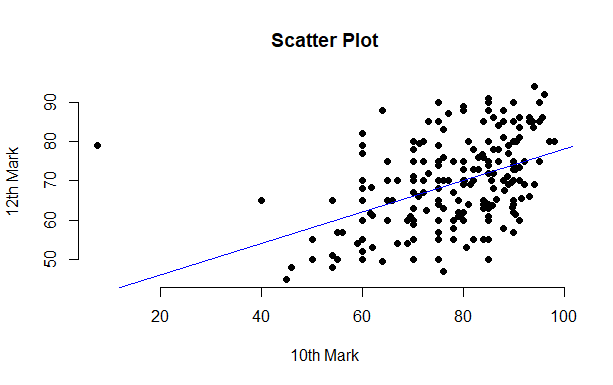
As can be seen in the result, the degree of freedom obtained is **265** while the t-critical value is **-2.0419**. In the t-table, we shall find its test statistics which in this case, the test statistics is **1.969** while the t-critical value is -2.0419. The critical value is smaller than the test statistics value. Thus, we can prove that both of the mean for both of the variables are different to each other, thus the null hypothesis, H0 can be rejected as it has sufficient evidence.

**Analysis:**

Correlation Analysis

**Result:**

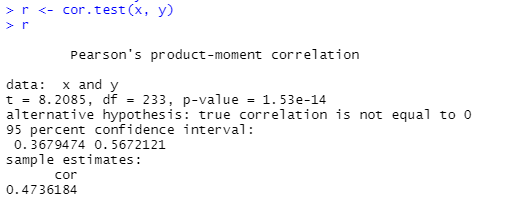
Diagram .. shows the scatter plot of the relationship between these chosen variables (12th Mark vs 10th Mark).



**Diagram**

From this graph, we can see that there is a positive **linear correlation between** 10th mark and 12th mark. The larger the 10th mark, the higher the 12th mark. So we can conclude that most of the students who got high marks in 10th level also got high marks in their 12th level.

From the calculation of correlation, we get the value of **r is 0.4736184** which indicates that there is a relatively **weak positive linear correlation** between x and y.

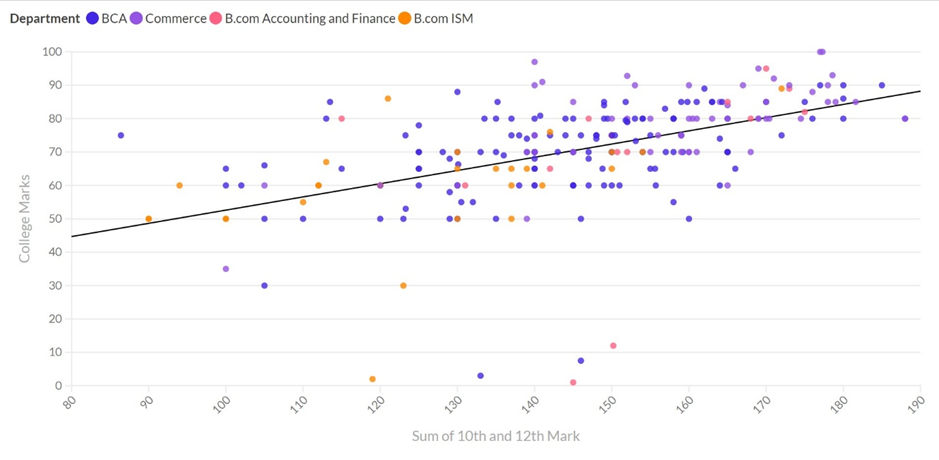


**Diagram**

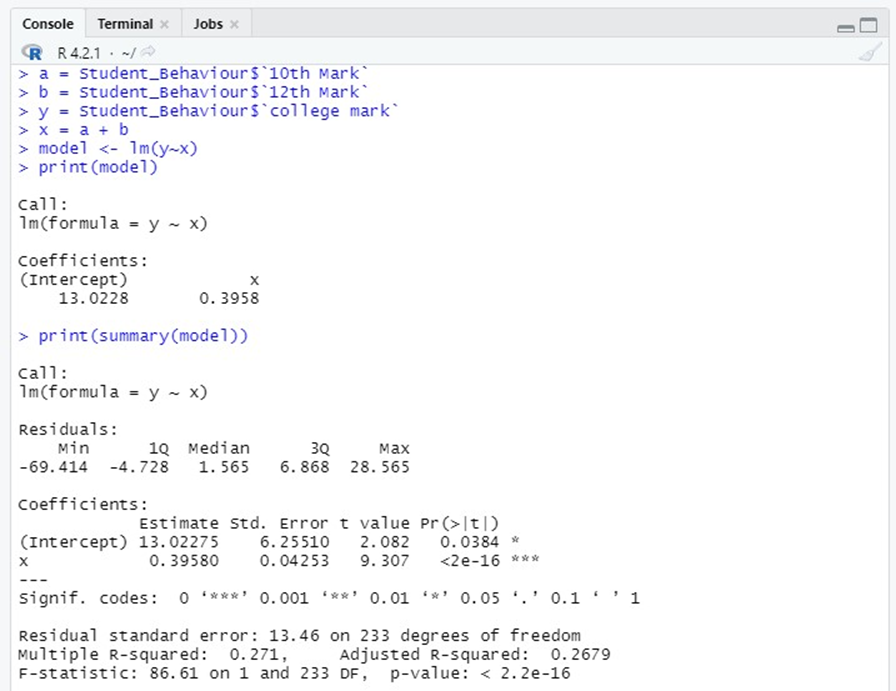
**Analysis:**

Regression Analysis

**Result:**

****

**Figure A: Scatter plot of College Marks against The Sum of 10th and 12th Marks**

****

**Figure B: Summary of Graph using RStudio**

*From figure B, we can get the the formula for estimated regression model is:*

**Y= 13.0228 + 0.3958X2x**

b0 is the estimated sum value of y when the value of x is zero

b1 is the estimated change in the sum value of y as a result of a one-unit change in x

From this, we can interpret that when the sum of 10th and 12th marks is **0**, the value of college marks is **13.0228**. From this, we can interpret that the value of college marks increases by **0.3958** as a result of each addition of the sum of 10th and 12th marks.

From Figure B, we can also get the coefficient of determination, is **0.2679**. This shows that only **2.679%** of the value of college marks is explained by the sum of 10th and 12th marks. This also shows that only some but not all the variation in the value of college marks is explained by variation in the sum of 10th and 12th marks.

**Analysis:**

Chi-Square Test Analysis

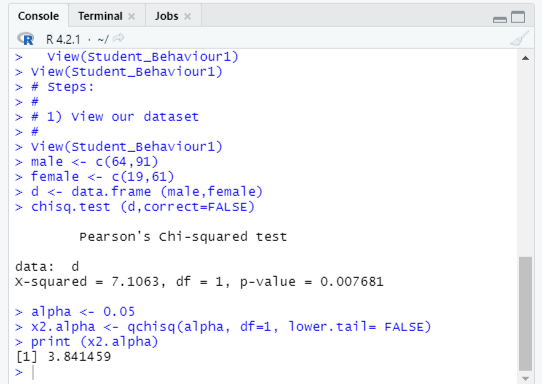
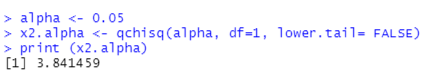
**Result:**

Figure above shows the overall R script to perform Chi-Square Test of Independence.

Above figure is the R script segment to determine statistics value χ2 , degree of freedom, df and p-value. From the figure above, we can find that the statistics value χ2 =7.1063, degree of freedom, df =1 and p-value= 0.007681.



Above Figure is an R script segment to determine the critical value χ2 = 3.841459. Hence, from this value, we can come to a conclusion that there exists a relationship between the two variables because statistics value is greater than the critical value. We **reject the null** hypothesis.

.

**9.0 WORK COORDINATION**

| **Group Members** | **Task Specification** |
| --- | --- |
| Aisyah | * Methodology * Descriptive Statistics * Inferential Statistics: 2-sample Hypothesis Testing Analysis * Interpretation of Results: 2-sample Hypothesis Testing Analysis |
| Immal | * Introduction * Scope & Objective * Inferential Statistics: Correlation Analysis * Interpretation of Results: Correlation Analysis |
| Mirza | * Dataset used * Inferential Statistics: Chi-Square Test Analysis * Reflections * Conclusion & Discussion |
| Izat | * Inferential Statistics: Regression Analysis * Statistical Data Analysis |

**10.0 REFLECTIONS**

Along the completion of this project, each of the group members have portrayed a very cooperative atmosphere. Every group member communicated efficiently despite being in different locations. By using Google Meet and WhatsApp, it is easy for us to apply our team working skills to complete this project. Not only that, we believe we have successfully applied all the necessary skills that have been taught by our lecturer, Dr. Azurah A Samah during lectures to perform test analysis. This test analysis refers to 2-sample hypothesis testing, Correlation testing, Regression testing and Chi-square test of independence. We also have sharpened our skills of using R language to perform analysis tests and mathematical problems in general.

Most importantly, we would like to express our utmost gratitude to our lecturer, Dr. Azurah A Samah for her guidance along the completion of this project. Our lecturer had given us ample time and flexibility to complete this project based on what we learned in lectures. To conclude, we strongly believe that this project will directly benefit us in our careers in the future.

**11.0 CONCLUSION & DISCUSSION**

11.1 Summarise Findings of Results

For the 2-Sample Hypothesis Testing Analysis, the test itself is simply comparing the mean between the two variables. Since we are comparing the mean of the college mark obtained by the students and daily time spent for studying, we can say that the higher the time spent for studying, the higher the college mark can be received.

For the Regression Analysis, if we take a closer look at the obtained P-value provided in the RStudio, we can understandably discern that the value itself is considerably lower than the significant level, α. Accordingly, we reject the null hypothesis, which leads us to believe that there is sufficient evidence to support that linear regression does indeed exist between the value of college marks and the sum of 10th and 12th marks. This also means that the sum of 10th and 12th marks have a transparent link with the value of college marks, giving us the message that they are both related in one way or another. Because of that, students need to be diligent to such an extent that both of the corresponding variables end in a flawless consequence.

For correlation analysis, we can conclude that there is a weak linear relationship between the 10th students’ mark and 12th students’ mark. Hence, we reject the null hypothesis as we have enough evidence to conclude that 10th mark relate with 12th mark of students. This shows that the good starting in academic life will give big impact on future study as students need to maintain their academic record due to high demanding on job application in the future.

For the Chi-Square Test for Independence, our study finds that there exists a relationship between Gender and Degree Completion of Certificate. Hence we **reject** the null hypothesis. In a real world scenario, gender might affect whether an individual wishes to complete any certificate based on their interests. This is because generally, females and males do not always share the same interest.

11.2 Discuss whether all objectives outlined in Section 2 have been achieved

Upon completion of this project, all the scopes and objectives that we have stated have been successfully achieved by each member of the group. By using Kaggle.com, we managed to perform secondary data collection as well as understanding the nature of the study hence determining all the necessary variables to use for this project. Secondly, we managed to organise all the datas and performed necessary modifications to change the variables to fill the requirements of Rstudio. Then, we applied suitable statistical methods to analyse our datasets. We also used an R statistical software (In this project, Rstudio) to summarise our variables. We then analysed patterns of interests of datasets by applying the suitable hypothesis testings. After that, each of the group members were able to interpret each of their results in order to make their results more significant. Lastly, our project is synthesised and documented individually in our e-portfolios. With that said, we can conclude that all objectives have been achieved together.

11.3 List of challenges faced throughout the study.

First and foremost, we faced challenges in terms of availability of each group members. The heavy workload and personal affairs that we have to manage while completing this project has made it difficult to have proper discussions. Secondly, we faced distance problems. Since not all of us are in campus, it is a bit difficult to have face-to-face meetings, hence there are communication problems between group members. Lastly, finding an appropriate dataset with suitable amount of variables and respondents was quite a challenge. This is due to vast amount of datasets from open sources made it difficult for us to find the best dataset.

11.4 Present improvement to be made in future

Surely, there are improvements that we can make to produce a better project. For example, we can spend more time and communicate much more efficiently to choose a more significant and interesting dataset. Secondly, we need to better understand the nature of the datasets so we can produce better hypothesis tests results. Lastly, time management is an aspect that we can surely improve. Despite having enough time to complete this project, we find it difficult to complete this project in time.

**12.0 REFERENCES**

12.1 Sources Of Data:

* <https://www.kaggle.com/dataset/8bde35079db4807b9a5b8e5bfc6c864e5f5b3143544af7a2b67bdd032b1fab59>

12.2 List Of Articles & References Used:

* <https://www.r-project.org/about.html>
* [https://www.statmethods.net/stats/descriptives.html#:~:text=R%20provides%20a%20wide%20range,with%20a%20specified%20summary%20statistic.&text=Possible%20functions%20used%20in%20sapply,median%2C%20range%2C%20and%20quantile.](https://www.r-project.org/about.html)
* [https://www.geeksforgeeks.org/regression-analysis-in-r-programming/#:~:text=Regression%20analysis%20is%20a%20group,independent%20variables%20of%20the%20dataset.](https://www.r-project.org/about.html)
* [https://www.datavedas.com/inferential-statistics-in-r/#:~:text=Inferential%20statistics%20are%20used,and%20predictions%20about%20the%20population.](https://www.r-project.org/about.html)
* <http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r>
* <https://www.statology.org/t-critical-value-r/>
* <https://www.datanovia.com/en/lessons/how-to-do-a-t-test-in-r-calculation-and-reporting/>
* <https://www.mastersindatascience.org/data-scientist-skills/r/#:~:text=As%20a%20programming%20language%2C%20R,data%20analysis%20and%20statistical%20modeling.>

12.3 List Of E-Portfolios:

* Aisyah: <https://eportfolio.utm.my/user/aisyah-binti-mohd-nadzri/seci2143-probability-statistical-data-analysis>
* Immal: <https://eportfolio.utm.my/view/view.php?t=nSsx8DNmMXI9wOE5Rfq4>
* Mirza:[AHMAD MIRZA ARMAND BIN SHAZRIL FARIZA - MyePortfolio@UTM](https://eportfolio.utm.my/user/ahmad-mirza-armand-bin-shazril)
* Izat: <https://eportfolio.utm.my/view/view.php?t=fBD1lwmsyFvQuU8pExOg>

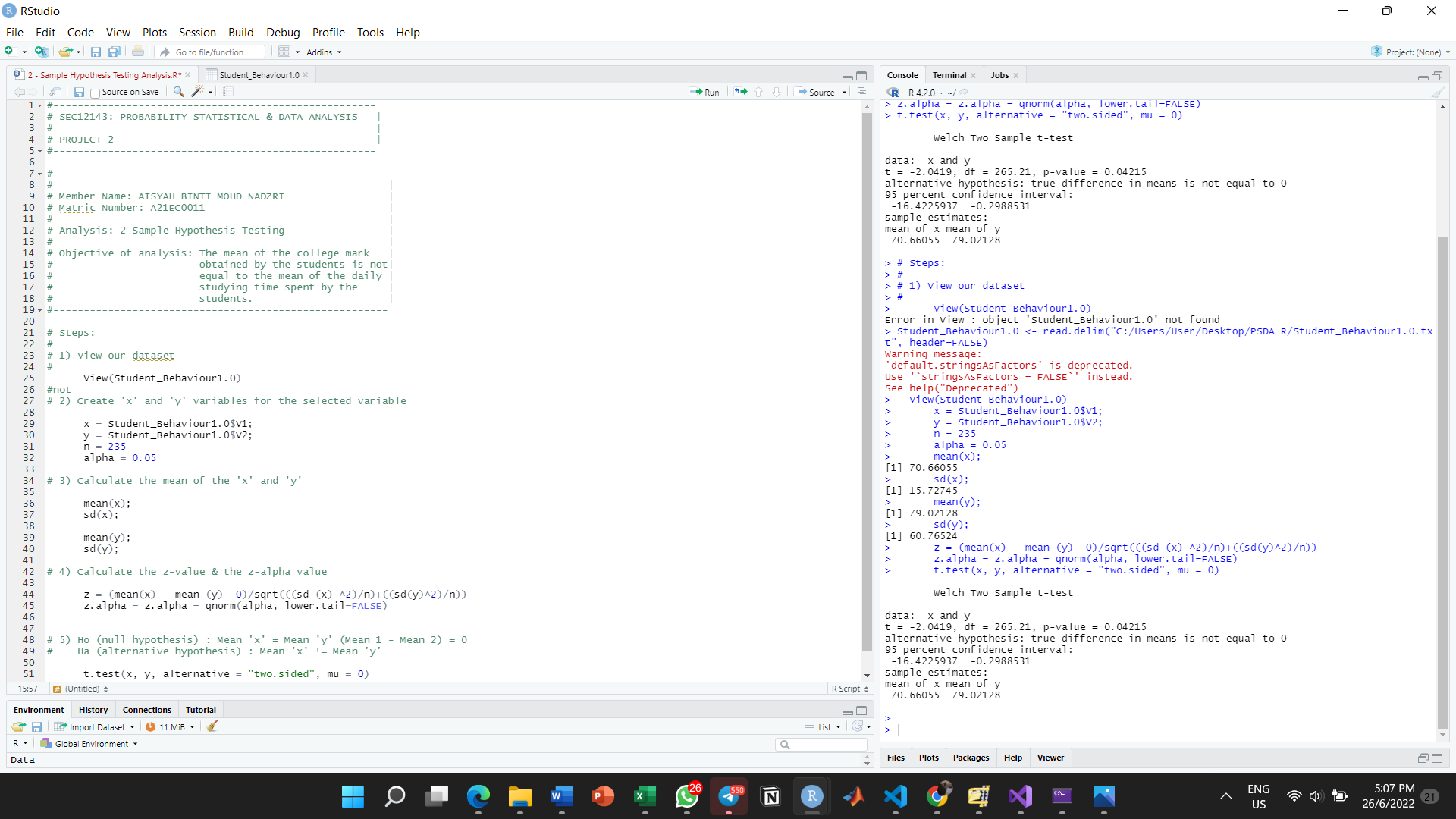
12.4 Link Of Presentation Video:

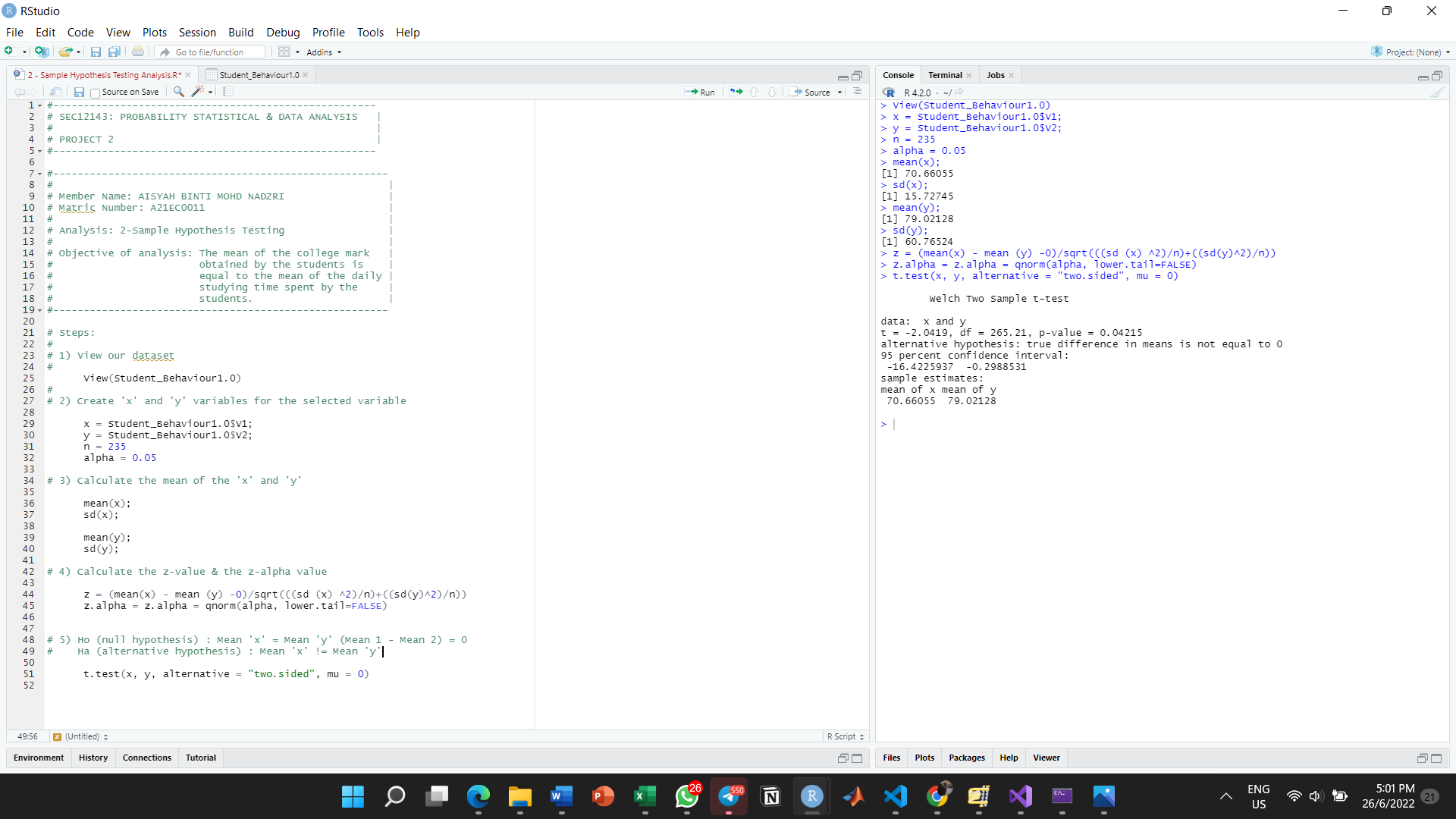
* <https://youtu.be/lVZEB0zmH3s>

**13.0 APPENDICES**

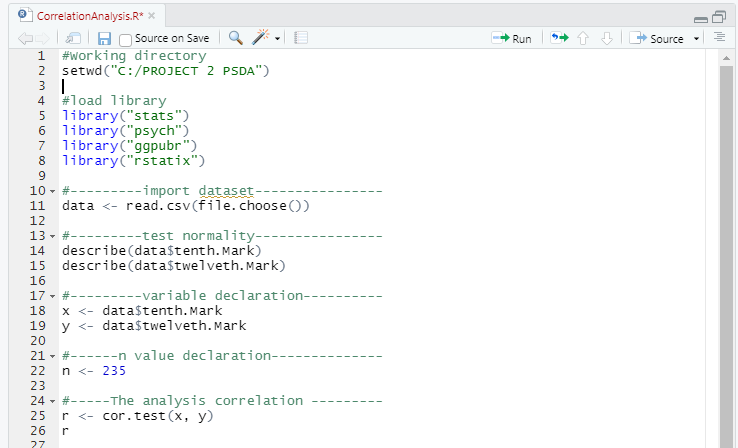
* Screenshot of Statistical codes:

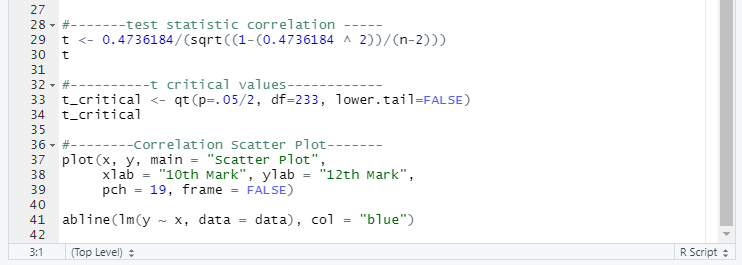
Analysis: 2-Sample Hypothesis Testing Analysis

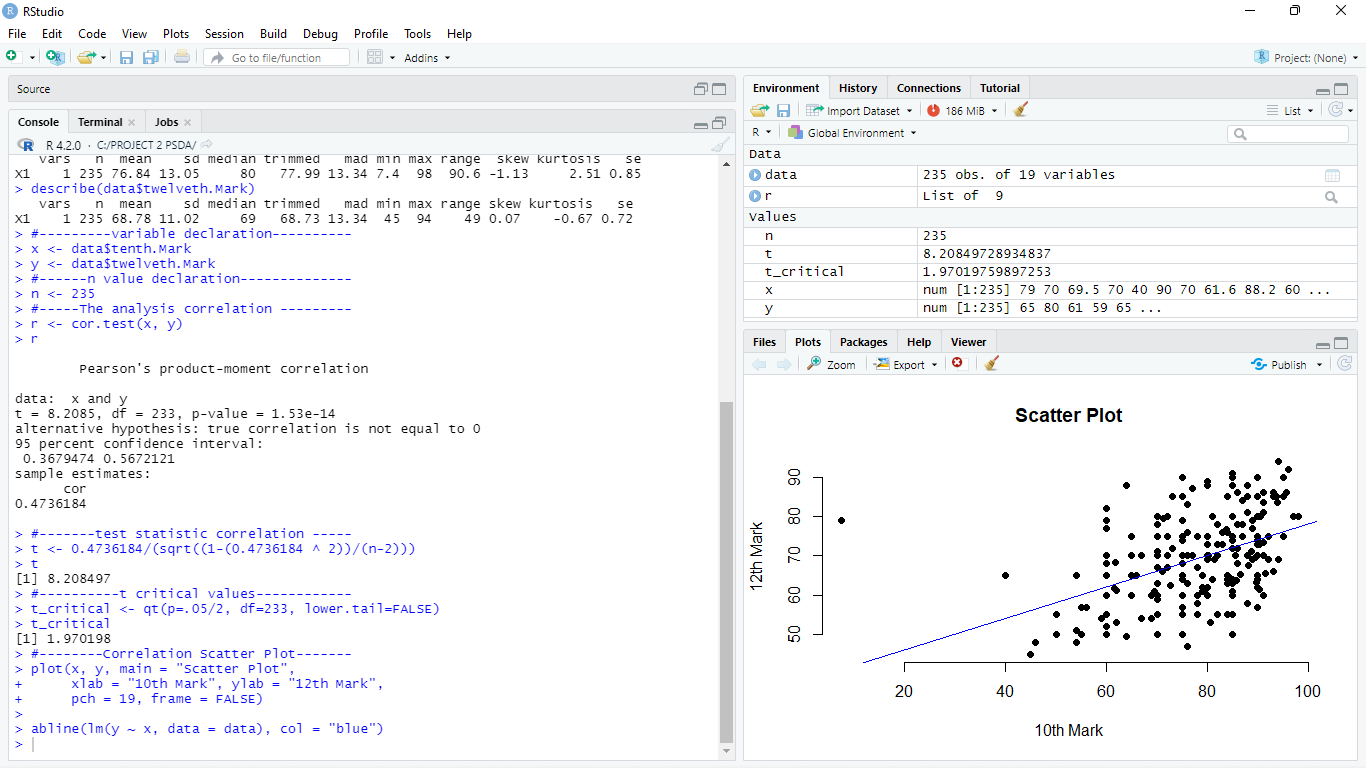
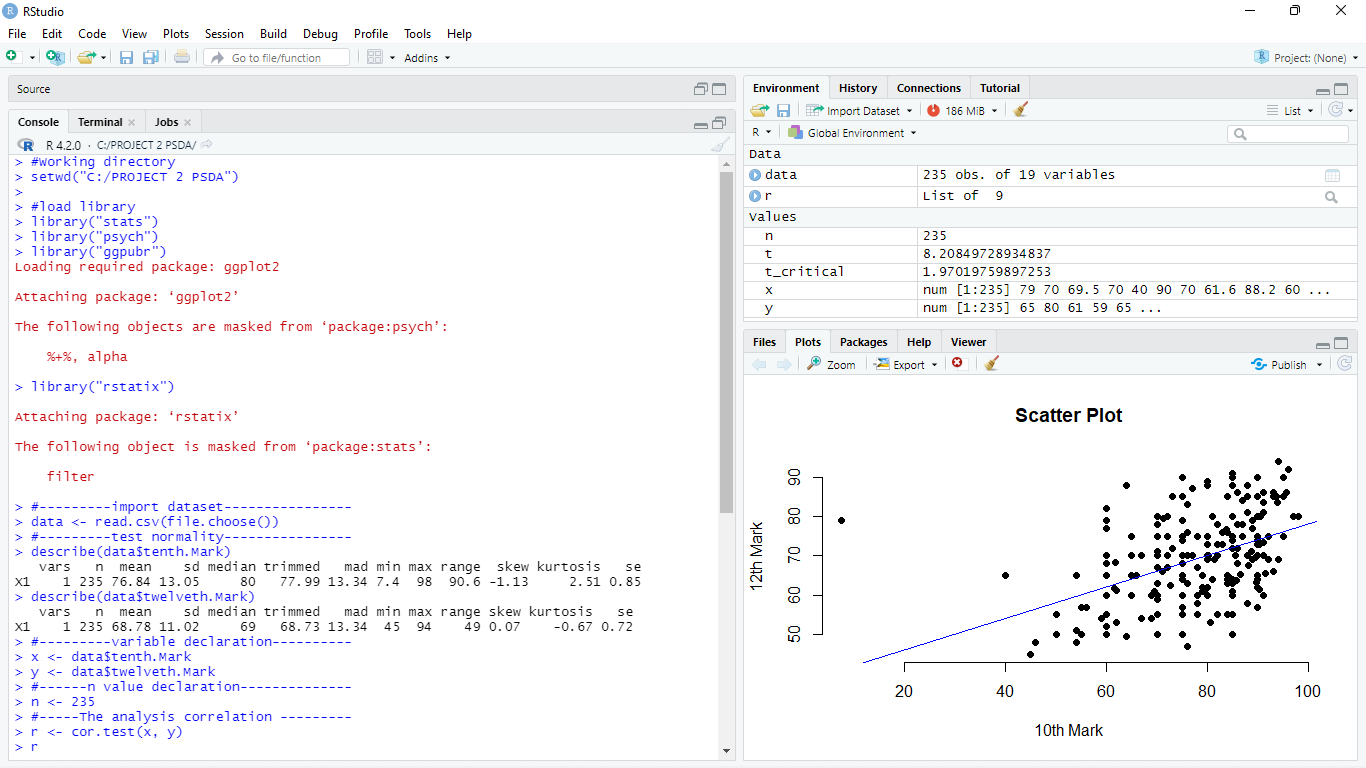




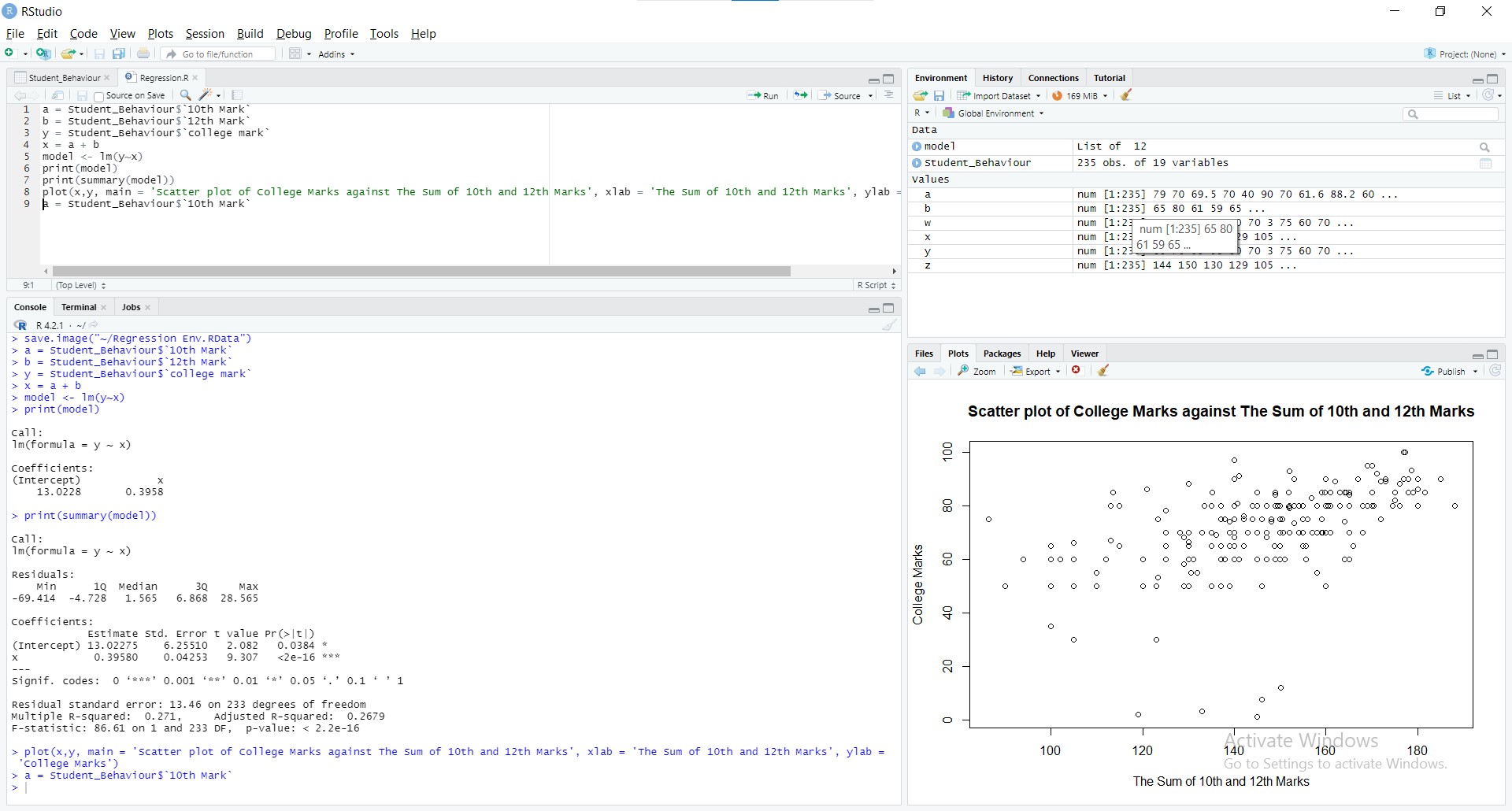
Analysis : Correlation Analysis

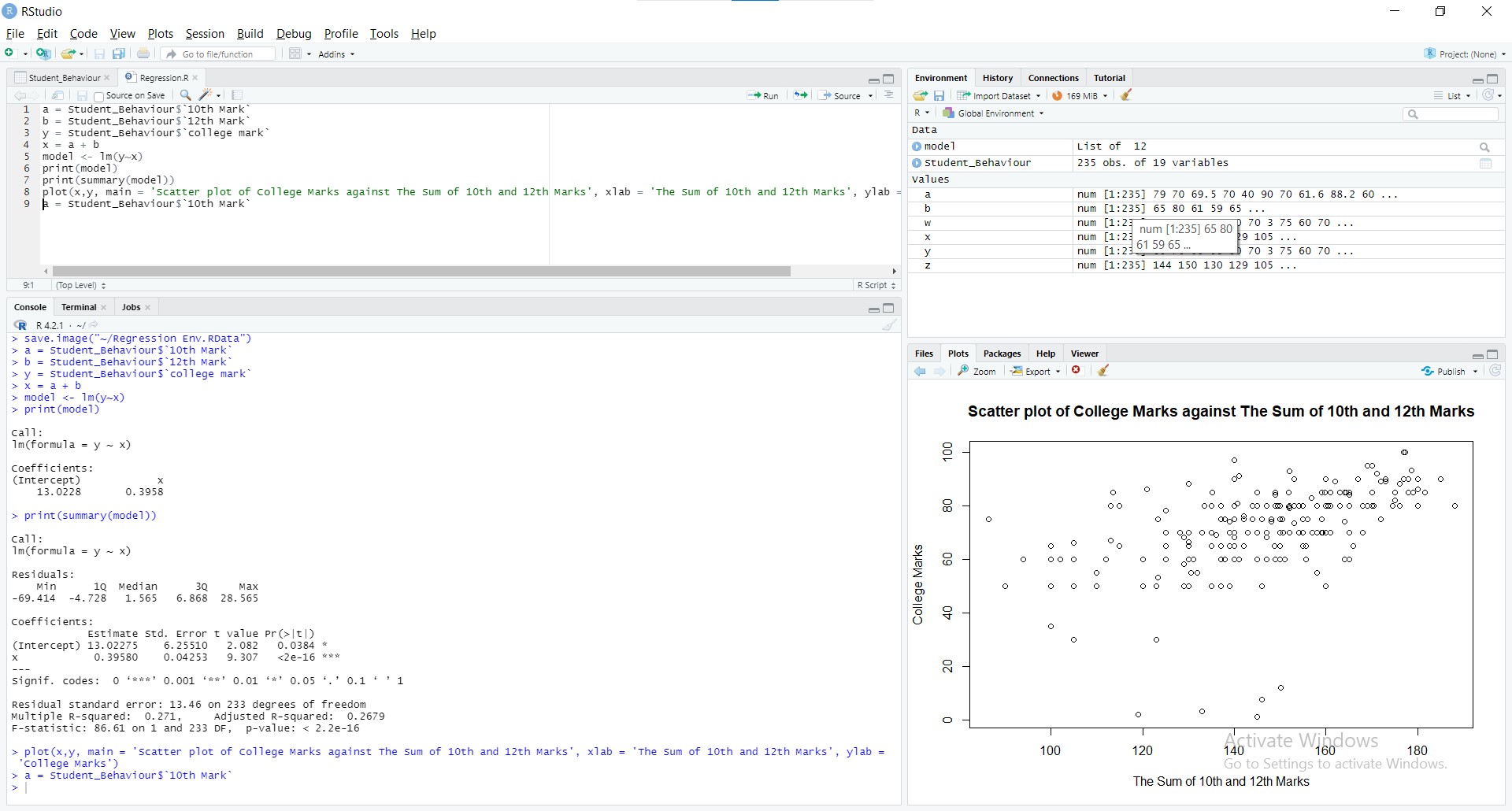




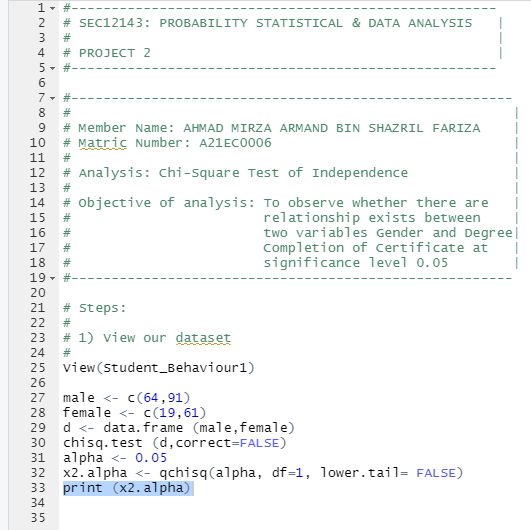


Analysis: Regression Analysis





Analysis: Chi-Square Test of Independence



* Photos of meetings & Discussion:

