



**SCHOOL OF COMPUTING**

**SEMESTER 2 2020/2021**

**PROBABILITY & STATISTICAL DATA ANALYSIS  
(SECI 2143-09)**

**TITLE: THE WEIGHT, INCHES AND RAM OF LAPTOPS OF THREE BRANDS**

**PROJECT 2**

**AMONG UTM:**

<b>NO</b>	<b>NAME</b>	<b>MATRIC NUMBER</b>
1.	NAZIHAH BINTI SURATI	A20EC5016
2.	SYAZA SYAURAH BINTI MOHD YUSRAN	A20EC0227
3.	DARRSHAN A/L RAJAGOPAL	A20EC0188

# Table of Contents

<b>1.0 Introduction.....</b>	<b>2</b>
<b>2.0 Dataset.....</b>	<b>2</b>
<b>3.0 Data Analysis.....</b>	<b>2</b>
<b>3.1 Hypothesis testing 2 samples.....</b>	<b>2</b>
<b>3.2 Correlation Analysis.....</b>	<b>3-4</b>
<b>3.3 Regression Analysis.....</b>	<b>4-5</b>
<b>3.4 ANOVA.....</b>	<b>5-6</b>
<b>4.0 Conclusion.....</b>	<b>6</b>
<b>5.0 References.....</b>	<b>6</b>

## **1.0 Introduction**

The topic of this project is the weight, inches, and ram of laptops of three brands. The aim of study is to investigate whether there is a relationship between the weight, inches, and ram of laptops of three brands which are Apple, Toshiba and MSI. Several statistical methods are used to estimate the relationship between the each of the variables. Statistics represent an essential part of a study because, regardless of the study design, investigators need to summarize the collected information for interpretation and presentation to others. The purpose of this study is to help create a data analysis plan for a quantitative study. Well, here on inference statistical analysis, it helps you go to conclusions and make predictions based on your data. When you have collected data from a sample, you can use inference statistics to understand the larger population from which the sample is taken. This is the reason why we are so interested into the question as well as knowing the statistical tests such as hypothesis testing, correlation, regression etc. So, what does the statistical test do? Statistical tests work by calculating a test statistic, a number that describes how much the relationship between variables in your test differs from the null hypothesis of no relationship. It then calculates a p-value (probability value). The p-value estimates how most likely it is that you would see the difference described by the test statistic if the null hypothesis of no relationship were true. If the value of the test statistic is more immense than the statistic calculated from the null hypothesis, then you can form an opinion that a statistically significant relationship between the predictor and outcome variables. Statistical tests can be parametric or non-parametric. Parametric tests are considered more statistically better because they are more likely to detect an effect if one exists. Based on the previous statement, that is what we are kind of expecting to see from the data.

## **2.0 Data set**

The data set used are obtained from online source. A population of 1303 data of laptops is chosen. Then, 123 data are randomly picked from the population, which is Apple, Toshiba and MSI. The data included type of laptops, inches, screen resolution, CPU, RAM, memory, GPU, OS, weight, and prices. A few variables are picked from the list for testing purpose which are type of laptops, weight, inches, and ram. Then, hypothesis testing, correlation, regression, and ANOVA are used to test the sample data. The ways of analysis of data are by using the RStudio to generate graphical presentation of data and to do some of the basic calculation. Then, the conclusions are drawn.

## **3.0 Data analysis**

### **3.1 Hypothesis testing 2 samples.**

Based on the data obtained, we want to study whether the weight of Apple Toshiba is same as the weight of MSI. So, hypothesis test for 2 independent sample are carried out. The population variances are assumed to be unknown. In this case, since the amount of data in the sample is 60 and 70 respectively. The weight of Apple Toshiba laptop is different from weight MSI laptop. Hence, the null hypothesis,  $H_0$  and alternative hypothesis,  $H_1$  is:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

where  $\mu_1$  represent population mean weight of Apple Toshiba,  $\mu_2$  represent population mean weight of MSI. The sample mean weight of Apple and Toshiba laptop,  $\bar{X}_1 = 1.5429$ , while the sample mean weight of MSI laptop is  $\bar{X}_2 = 2.6606$ .

The sample standard deviation is calculated using formula.

$$s = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}}$$

for both of the samples. So, the result are  $S_1 = 0.4423$  and  $S_2 = 0.6939$ .

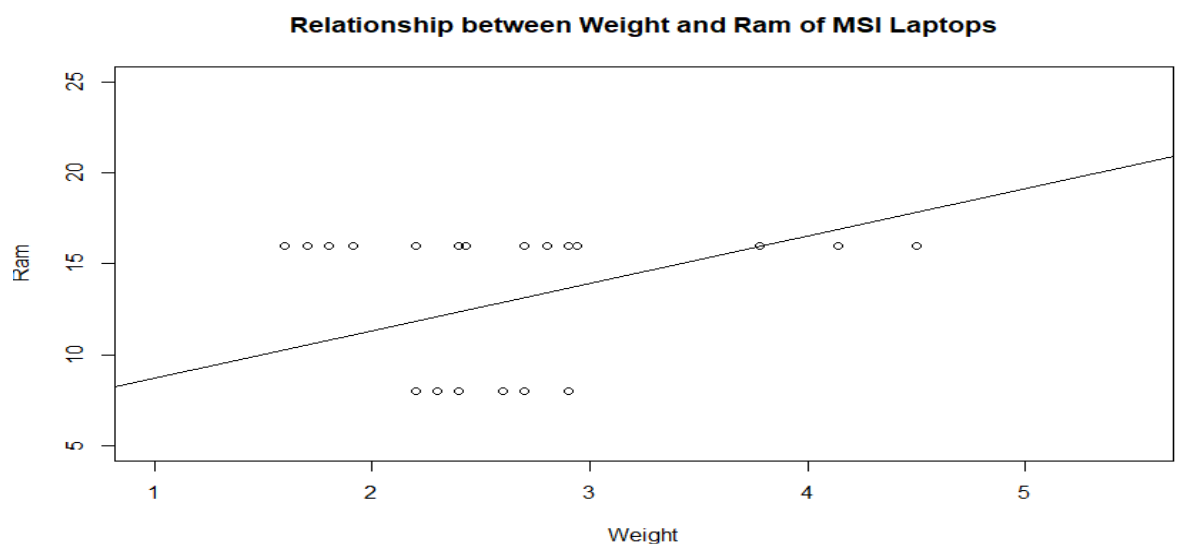
Since, it is two-tail test, the critical value of 0.05 significance level is  $\pm 1.988268$ .

Conclusion: We failed to reject  $H_0$ , since  $-1.988268 > -10.31039 < 1.988268$ . We can conclude that we have strong evidence that the weight of Apple and Toshiba is same weight as MSI laptop.

### 3.2 Correlation

In the correlation test, we had analysed that the relationship between the weight of MSI and the ram of MSI laptops in the sample of 54. The coefficient correlation,  $r$  is calculated using SPSS to show the relationship between these two variables. We use Pearson's technique to calculate since both data are ratio type. These two variables are used to calculate the coefficient correlation using the formula:

$$r = \frac{\sum xy - (\sum x \sum y)/n}{[(\sum x^2) - (\sum x)^2/n][(\sum y^2) - (\sum y)^2/n]}$$



```

Pearson's product-moment correlation

data:  x and y
t = 2.9768, df = 52, p-value = 0.004414
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1267616 0.5891393
sample estimates:
      cor
0.3815709

```

The correlation coefficient,  $r = 0.3815709$ . A scatter plot and correlation coefficient,  $r$  indicates that there is positive relationship between the weight and ram of MSI laptops.

### **Significance Test for correlation**

Then, these two variables are used to test whether there is any evidence of linear relationship between the weight and ram of MSI laptops at the 0.05 level of significance.

$H_0: \rho = 0$  (no linear correlation)

$H_1: \rho \neq 0$  (linear correlation exists)  $\alpha$

In this test, the test statistics,  $t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$ ,  $t$  is 2.9768. P-value is the significance level of the test t-test. P-value is 0.004414. Confident interval of the correlation coefficient at 95% is (0.1267616 ,0.5891393).

Conclusion: We reject  $H_0$ , since  $0.004414 < 0.1267616$ . There is sufficient evidence of a linear relationship between the weight and ram of MSI laptop at the 0.05 significance level.

### **3.3 Regression**

In this regression test, we measure the relationship between the weight and ram of MSI laptops in the sample size of 54. The independent variable is weight, and the dependent variable is ram. The sample regression line provides an estimate of the population regression line.

Estimated Regression Model:

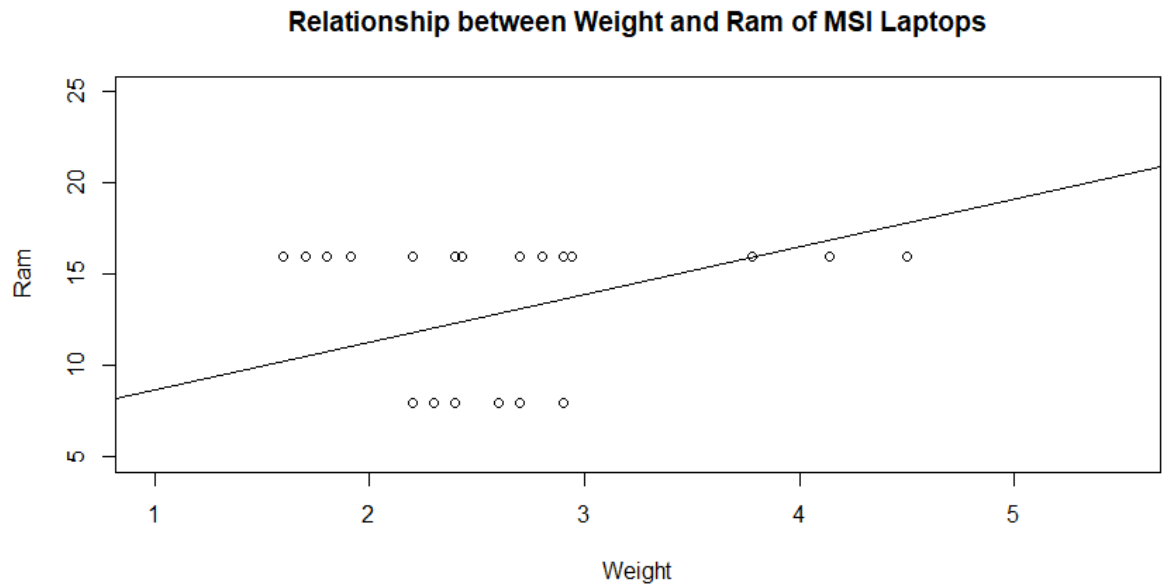
$$Y = b_0 + b_1X$$

Where,  $y$  is dependent variable.

$b_0$  is the population intercept  $y$ .

$b_1$  is the population slope coefficient.

$x$  is independent variable.



```
Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-5.661 -3.837  0.046   3.564 14.431

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.1045     2.4054   2.538  0.01419 *
x              2.6057     0.8753   2.977  0.00441 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.422 on 52 degrees of freedom
Multiple R-squared:  0.1456,    Adjusted R-squared:  0.1292
F-statistic: 8.861 on 1 and 52 DF,  p-value: 0.004414
```

The estimated regression model is then calculated using RStudio, where we obtain  $\hat{y} = 6.105 + 2.606x$

### **Test Statistical of Regression**

$H_0: \beta_1 = 0$  (no linear relationship)

$H_1: \beta_1 \neq 0$  (linear relationship)

Test statistic,  $t = \frac{b_1 - \beta_1}{Sb_1}$

Where,  $b_1$  = Sample regression slope coefficient

$\beta_1$  = Hypothesized slope

$Sb_1$  = Estimator of the standard error of the slope

Conclusion: We reject  $H_0$ , since 0.004414 is less than significance level 0.05. So, there is sufficient evidence that the weight of MSI laptops affects the ram of MSI laptops.

### 3.4 ANOVA

In this test is to test for significance differences between means of inches among 3 brands of laptops at the 0.05 significance level.

$H_0: \mu_1 = \mu_2 = \mu_3$  (all brands have the same mean of inches)

$H_1$ : at least one mean is different.

$$F = \frac{\text{variance between samples}}{\text{variance within samples}}$$

```
      Df Sum Sq Mean Sq F value Pr(>F)
brand    2  196.1    98.03   86.2 <2e-16 ***
Residuals 120  136.5     1.14
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion: We reject  $H_0$ , since P-value of  $F = 2 \times 10^{-16} = 0.00000000000000010$  is less than the significance level 0.05. So, there is insufficient evidence to claim that the different types of laptops among 3 brands have the same mean of inches.

### 4.0 Conclusion

After all the insights we got from the statistical analysis we did, from the hypothesis test, we can conclude that the weight of Apple and Toshiba laptop is same weight as MSI laptop. From the correlation test, it shows the result that the weight and ram of MSI laptops have a positive relationship, that is if the laptop is heavier, it has larger ram. From the regression test, for MSI laptop, it shows that for each increase of 1kg of weight of laptop, the ram also increased by 2.606. From the ANOVA test, it shows that the mean of inches of each brand is different.

Throughout the project, it was very interesting because after we analysed the variables or issues that we want to analyse, we can know more about it and determine it is true or not and are there dependent and even their relationship between each other. By doing this project, we also got to strengthen our relationship as teammates in order to deliver the best of work for this project. We are exposed to new knowledge in using RStudio to calculate the research data. The project really gives us so much information regarding laptops and its variables with their own brands. Therefore, we can say that this helps us a lot in doing research and be more proactive in using new development in order to get the task done.

### 5.0 References

<https://www.kaggle.com/ionaskel/laptop-prices>