SECI2143 PROBABILITY AND STATISTICAL DATA ANALYSIS
PROJECT II – STATISTICAL ANALYSIS

# Relationship between time management and academic performance

Ahmedalla Hanea Hamza Hafez Hanoura
Section 07
School of Computing, Faculty of Engineering
University Technology Malaysia

***Abstract-*** *This case study is made to know if there is a relationship between time management and academic performance, and how can they affect each other. It is very important to conduct this case study because it is describing the importance of time management that affects students' performance and attitude.*

## I. INTRODUCTION

Time management is the way toward arranging and practicing cognizant control of time spent on explicit exercises, particularly to build adequacy, proficiency, and profitability. University students these days continuously commented that they don't have sufficient time to total all the missions delegated to them, Time administration is exceptionally critical and it may influence an individual's general execution and accomplishments. In fact, time management is very essential in all aspects of life, and many people rely on it constantly. Due to that, the objective of this project is to determine if there is a relationship between time management and academic performance in many aspects, according to the data given collected from Nottingham Trent International College students.

## II. METHODOLOGY

I collected my data using kaggle.com which is a very famous and known website when it comes to researches and data collection. Data were collected by DarwinLi, by a survey, that was answered by students of Nottingham Trent Worldwide College. I will use R studio also to help me conduct my statistical tests in this project such as, hypothesis test (2 sample), correlation, regression, Chi Square test of independence, and Goodness of fit test.

## III. RESULTS AND DISCUSSION

Hypothesis testing (2 sample)

Two sample test on level of organizing time issues among students between gender, I wish to determine whether there is any difference between the mean scores between males and females when it comes to organizing time issues. The number of chosen male students is 50, while female students is 45. I will assume that the variances of both genders are not equal since we can't reasonably predict that both of the data are having the same variances.

Let $\mu_1$ = sample mean of male in organizing time issues.

Let $\mu_2$ = sample mean of female in organizing time issues.

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Significance level, $\alpha = 0.05$

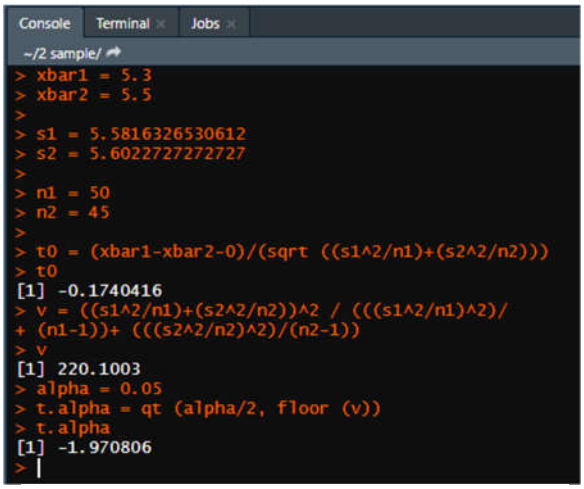
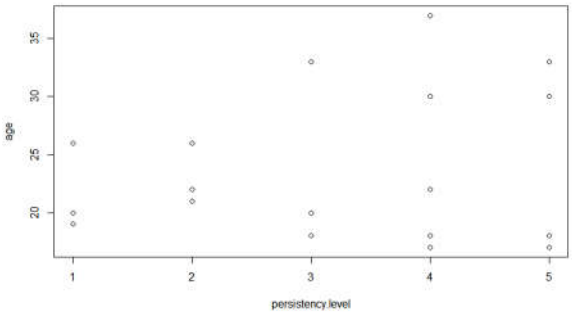



*Figure 1: R calculation for hypothesis test*

Null hypothesis ($H_0$) will be rejected if $t_0 < -t_{0.05,220}$ or if $t_0 > t_{0.05,220}$.
We can see that ($t_0$ = -0.174) > ($-t_{0.05,220}$ = -1.971) and ($t_0$ = -0.174) < ($t_{0.05,220}$ = 1.971). So, we fail to reject $H_0$ because it didn't fall in the critical region. There is an evidence to that the mean of male in organizing time issues is not different from the mean female in organizing time issues. Hence, we can conclude that the sample mean of male and female are the same when it comes to organizing time.




*Graph 1: Scatter plot of persistency level and age*

Correlation

I did select a sample of a 20 student for this test. In this test we will have the age of students and the persistency level for each one to know if there a linear correlation between them or not.

$H_0$ : There is no linear correlation between persistency level and the age of students.

$H_1$ : The linear correlation exists between persistency level and the age of students.

```
Console   Terminal   Jobs
C:/Users/xcite/Desktop/
> persistency.level = c(4, 1, 3, 4, 5, 4, 5, 3, 4,
+ 2 ,4 ,5 ,1 ,2, 4, 5, 3, 2, 1, 5)
>
> age = c( 17, 19, 18, 30, 33, 37, 18, 20, 22, 22,
+ 18, 18, 26, 21, 17, 17, 33, 26, 20, 30)
>
> cor (persistency.level,age)
[1] 0.07186928
>
```

*Figure 2: R calculation for correlation test*

Based on the calculated data above, we can see that the result of correlation between age and persistency level is equal to (0.07186928), Thus we can conclude that there is a very weak or even no relationship between age and persistency level. We can make sure that the relationship is very weak by referring to scatter plot that have been produced using R code in "Graph 1" we can see clearly our test result in it.

Regression

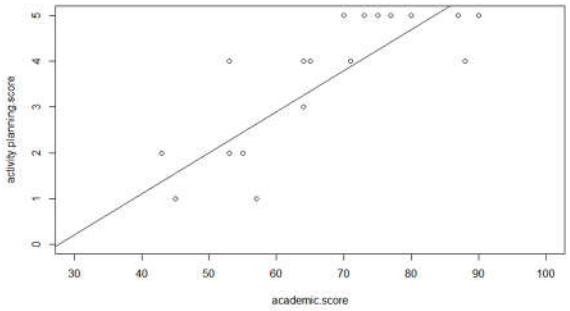
I selected A random sample of 20 students for this test. The dependent variable (y) in this case is the collected activity planning score, and the independent variable (x) is the students' academic score in university. This analysis aims to test the existence of a linear relationship between the variable x and y.

$H_0$: β1 = 0. (null hypothesis)

$H_1$: β1 ≠ 0. (alternative hypothesis)

```
Console   Terminal   Jobs
C:/Users/xcite/Desktop/
> activity.planning.score = c(5, 4, 5, 2, 1, 5, 5, 4, 4, 5, 5, 2,
+ 2, 1, 3, 4, 4, 1, 1, 5)
>
> academic.score = c(70, 65, 80, 53, 45, 75, 73, 71, 64, 87, 77,
+ 55, 43, 57, 64, 53, 88, 45, 57, 90)
>
> cor(activity.planning.score,academic.score)
[1] 0.8148127
>
> model = lm(activity.planning.score~academic.score)
> model

Call:
lm(formula = activity.planning.score ~ academic.score)

Coefficients:
    (Intercept)  academic.score
        -2.4843          0.0897

>
> summary(model)

Call:
lm(formula = activity.planning.score ~ academic.score)

Residuals:
     Min       1Q   Median       3Q      Max
-1.62859 -0.55220 -0.07043  0.67624  1.73021

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.48426    1.00960  -2.461   0.0242 *
academic.score  0.08970    0.01504   5.963 1.22e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9545 on 18 degrees of freedom
Multiple R-squared:  0.6639,    Adjusted R-squared:  0.6452
F-statistic: 35.56 on 1 and 18 DF,  p-value: 1.216e-05
```

*Figure 3: R calculation for regression*




*Graph 2: plot of planning score and academic score*

From the previous calculations and graph, we can see that model is in a positive linear relationship with equation of y = -2.4843 + 0.0897x. in this test the relationship between activity planning score and students' academic score is 0.8148 which is close to 1 that we can consider it as a strong relationship between the two variables, we can determine that if students plan their activities better they will achieve a higher grade or score in their academic performance. Actually, it very convincing because as we can see in the graph that students who score 5 in their activity planning, they achieved 70 or more when it comes to the academic score.

Goodness of fit test

This test is to find out whether there is a relationship between the age of students and their attendance rate.
Let:
$P_1$ = S0, $P_2$ = S1,
$P_3$ = S2, $P_4$ = S3, $P_5$ = S4

Significance level, α = 0.05

$H_0$: $P_1 = P_2 = P_3 = P_4 = P_5 = 1/5$.

$H_1$: At least one of the proportions is different from the others.

| | S0 | S1 | S2 | S3 | S4 |
|---|---|---|---|---|---|
| age<18 | 3 | 2 | 0 | 1 | 0 |
| 18-20 | 16 | 9 | 4 | 3 | 4 |
| 21-25 | 32 | 15 | 5 | 4 | 1 |
| 26-30 | 7 | 2 | 2 | 1 | 0 |
| 31-35 | 2 | 1 | 1 | 0 | 0 |
| 36>age | 1 | 1 | 1 | 1 | 0 |

*Table 1: Table of age range and attendance rate*

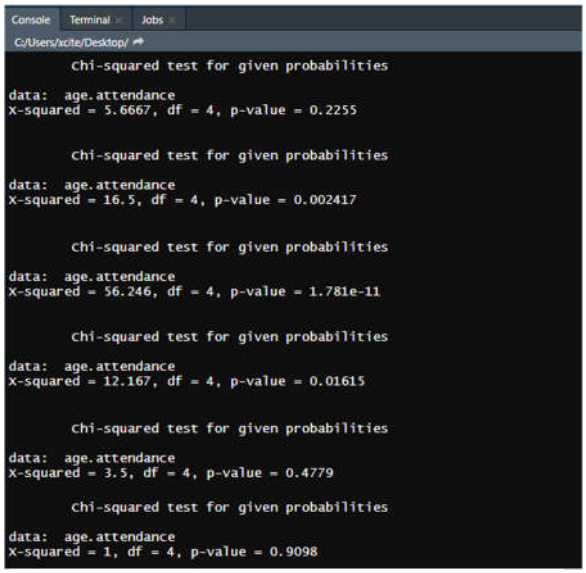



*Figure 4: R calculation for goodness of fit*

$x^2_{0.05,4} = 9.488$

As we can see from the calculated data above using R software, if the test statistic is > 9.488 we will reject $H_0$ but if it smaller we will fail to reject $H_0$. In this case from the age of 18 to 30 we reject $H_0$, and younger than 18 or older than 31 we fail to reject $H_0$. We can conclude distribution is not uniform for students between 18 to 30, and it is uniform for students who are younger than 18 and older than 31.

Chi square test of independence

This test is to find out whether there is a relationship between the attendance rate of students and their academic scores. I chose the same sample of 20 students from previous regression test for this test.

$H_0$ : The attendance rate is independent of the academic scores of students.
$H_1$: The attendance rate is dependent on the academic scores of students.

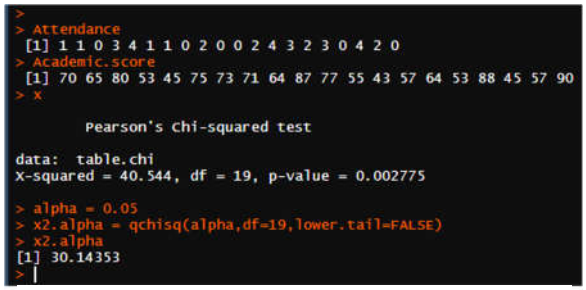



*Figure 5: R calculation for Chi square test*

```
> table.chi
          Attendance(S)       Academic score       Total
Student1              1                   70          71
Student2              1                   65          66
Student3              0                   80          80
Student4              3                   53          56
Student5              4                   45          49
Student6              1                   75          76
Student7              1                   73          74
Student8              0                   71          71
Student9              2                   64          66
Student10             0                   87          87
Student11             0                   77          77
Student12             2                   55          57
Student13             4                   43          47
Student14             3                   57          60
Student15             2                   64          66
Student16             3                   53          56
Student17             0                   88          88
Student18             4                   45          49
Student19             2                   57          59
Student20             0                   90          90
```

*Table 2: Table of academic score and attendance*

As we can see from the calculation and the table above, we can make some decisions but first I need to mention that when attendance(S) = 0 that means the most attendance and when it is equal to 4 that means the least attendance. From the calculated data we can determine that critical value is = 0.002775, and chi square value is = 40.544. Since the chi square value is more than the critical value, we reject the null hypothesis ($H_0$) and concluded that the attendance rate is dependent on the academic scores of students. So, there is a relationship between both attendance and academic scores and they rely on each other.

## IV. CONCLUSION

In conclusion, we can make a lot of decisions about this study. Based on the data from this study, a conclusion can be made. As a first point, both male and female students can organize time, and there is no difference in between them in this case. Secondly, there is a very weak relationship between the age of students and their persistency that means that any student can be persistence whatever his/her age is. Thirdly, there is a very strong relationship between activity planning and students academic scores the more time students spent while planning that means a better score to achieve. Fourthly, when it comes to attendance you can determine from this study that students older than 31 and younger than 18 years old is attending more than students between this range. Lastly, we determined that students who attend to their classes more they will get a better grade. Finally, I achieved my purposes in this study by knowing if time management is affecting academic score of students or not, and clearly, we can see that it does.