

SECI2143-10 PROBABILITY & STATISTICAL DATA ANALYSIS

PROJECT 2

PREPARED BY:

NURUL SYAMIRA BINTI AMAT JIFRI (A19EC0145)

PREPARED FOR:

DR. AZURAH ABU SAMAH

**TABLE OF CONTENTS**

Introduction…………………………………………………………………… 2

Hypothesis Testing

 One Sample T-Test…………………………………………………… 3

 Correlation & Regression…………………………………………….4-8

 Anova …………………………………………………………………9

Discussion and Conclusion……………………………………………………10

**INTRODUCTION**

A simulation environment named Deeds (Digital Electronics Education and Design Suite) which is used for e-learning in Digital Electronics. It was conducted on a group of 115 students of first-year, undergraduate Engineering major of the University of Genoa.

The environment provides learning materials through specialized browsers for the students, and asks them to solve various activities which regarding to the subject with different levels of difficulty, and also different session times to conduct the activities. Differently from the proposal that already been submitted, I had to change my purpose of study which to see does the time taken by students do affects the grades achieved.

|  |  |
| --- | --- |
| Variables | Type |
| Session (which laboratory session) | Nominal |
| Student ID | Nominal |
| Activity (Activity selection in topic) | Nominal |
| Time taken (in minutes) | Ratio |
| Grades (in points /100) | Interval |

For the activtivy list:

* The Digital Circuit Simulator (DCS)
* The Finite State Machine Simulator (FSM)
* The Microcomputer Board Emulator (MCE)

**HYPOTHESIS TESTING**

**ONE SAMPLE T-TEST**

Hypothesis

* Null hypothesis (**H0**): the coefficients are equal to zero (i.e., no relationship between *Time taken in minutes* (x) and *Grade(*y*)* )
* Alternative Hypothesis (H1): the coefficients are not equal to zero (i.e., there is some relationship between *Time taken in minutes* (x) and *Grade(*y*)*)

****

In the output above:

* Test statistic value, t = 20.741
* Degrees of freedom, df = 114
* The significance level of the t-test, p-value = 2.2e-16
* Confidence interval of the mean difference 95% is shown (48.93338, 59.16227)
* Sample estimates which is the mean differences is 54.04783.

Summary

From the ouput above, with the value of t = 20.741 > p-value = 2.2e-16, we fail to reject the null hypothesis (**H0**). There is enough evidence to supoort that the time taken by students to complete their task will not affect their grades.

**CORRELATION AND REGRESSION**

We create a scatterplot to study the correlation between the two variables. In figure below, there shown there is no relationship between the variable.

****

To properly quantify the relationship between the two variables, I calculate the correlation coefficient using a function called *cor* on Rstudio.



The value obtained is 0.1871265. As the value is closer to 0, the weaker the linear relationship. That shows the no relationship between two variables.

Here is the data analysis using linear regression model.

**VISUALIZATION**



The graph above suggest a no relationship between the *time taken* and the *grade* variables. This is a bit off since no regressin line can be made on the scatter plot.

**COMPUTATION**



**INTERPRETATION**

From the output above:

* The estimated regression line equation can be written as follow:

Data$‘Time taken(in minutes)’ = 98.42 + 0.083\*data$Grade

* The intercept (b0) is 98.42. It can be interpreted as the
* The regression beta coefficient for the variable Grade (b1), also known as the slope, is 0.083.

**REGRESSION LINE**

From the scatter plot that I did on Rstudio, it shows that between Time taken and Grades has no relationship.

**MODEL ASSESSMENT**

SUMMARY OF MODEL ASSESSMENT

The output of the six components are as follows:



Residuals

The median is -2.222 which is far from 0. It is negative so, they are below the regression line.

Co-effecients

* Null hypothesis (**H0**): the coefficients are equal to zero (i.e., no relationship between *Time taken in minutes* (x) and *Grade(*y*)* )
* Alternative Hypothesis (H1): the coefficients are not equal to zero (i.e., there is some relationship between *Time taken in minutes* (x) and *Grade(*y*)*)

Residuals Standard Error

From my model, RSE = 12.07 which is very far from 0. This shows a quite unacceptable prediction error.

R squared

In my dataset, the R-squared is 0.03502 which is quite low in value. This shows that a regression that does not explain the variance in the response variable well.

F-statistics

In my dataset, The F-statistics equal to 4.1 and producing p-value = 0.04523 is significant. This shows the value of 4.1 which is a little bit larger than 1 is already sufficient to reject the null hypothesis (**H0**).

p-value

p-value produced is 0.04523.

**ANOVA**

Anova is a method to test the significant difeerences between means.

For ANOVA, I used F-test to test the one-way ANOVA.

The statement hypothesis:

**H0**:μ1=μ2=μ3=μ4H1: at least one mean is different.



From the output above, the F-value is 4.1. Since F test statistic = 4.1 > F critical value, reject null hypothesis (**H0**).

**DISCUSSION AND CONCLUSION**

In conclusion, “Time taken (in minutes)” is not a very good predictor for the analysis. This is because in an examination, experiment or activities there are rarely the grades that will be achieved by students are effected by the time taken they completed a certain task. Logically, someone’s intelligence doesn’t be measured by how fast or active they are moving.