UNIVERSITI TEKNOLOGI MALAYSIA

JOHOR BAHRU (UTM JB),

81300 JOHOR BAHRU,

JOHOR

SCHOOL OF COMPUTING

FACULTY OF ENGINEERING

Project 2

Factors affecting Performance of Students

Name	Rakesh A/L Kannapathy	A19EC0153
Course Code & Name	SECI 2143 – Section 06 (SECJ) Probability and Statistical Data Analysis	
Lecturer	Dr.Chan Weng Howe	
Submission Date	2/7/2020	

Table of Content

Introduction

During the age of globalization, human beings are evolving themselves by learning new things. This is to ensure that new knowledge obtained will be used for the purpose of improving human lifestyle. This can be related to students all over the world which keep on learning new things. Each and one of them have different performance in their studies. There are a lot of factors that affects the performance of them. One of it is consumption of alcohol. Alcohol has been proven to have pros and cons that affects human. One of them is consuming alcohol can protect your heart. The National Institute on Alcohol Abuse and Alcoholism has confirmed that can reduce the risk of its consumer to get heart disease. A moderate consumption of it has been proven that it gives a lot of good effects to human. Other than that, consumption of alcohol will also give a bad impact to the user. One of them is increasing of crime rates as what are happening in our country, drunk driver involved in accident which can cause death and harm to other human beings. This also can cause bad impacts for us especially students who consume alcohol during their studies. The group of participants is being focus to teenagers which is an interesting group of focus. The statistical data that I obtained is about consumption of alcohol among students. Hence, based on the data, there are a lot of variable that effecting the student performance which will be explained.

The main purpose of this project is to study about 395 number of participants which consist group of focus students from multiple high school. The list of variables that I have used in this project is Gender, Grade 3, StudyTime, DailyAlcohol, Absence. The first test that I have done is to test whether type of gender has the same mean of grade 3. The second test that I done is to test whether there is relationship between alcohol consumption and study time. The third test is to observe the relationship between Absence and Grade3.Lastly, the test that I have done is to prove whether gender is independent with higher education.

Methodology

The type of data that I used in this project does contain both qualitative and quantitative data. All the calculation done in this project using quantitative data while qualitative data is based on critical analysis that has been included in this report.

The dataset that I have used in this project is called “Student Alcohol Consumption” that available in data.world website. This dataset is a secondary data as I obtained it from another user. The data sets contain 395 number of participants and contains a total number of 34 column of variable. Out of all the number of variables contains in the data, I have chosen 5 different variables to carry out the test which helps me to do critical analysis based on these variables that I have chosen. The type of analysis that I used in this project is Hypothesis Testing 2 Samples, Correlation using Spearman Correlation coefficient, Regression and Chi-Square Test two contingency that is independence test. All the calculation and analysis that have been done does use a software called RStudio.

Data and Analysis

1.Hypothesis Testing 2-Samples Mean

The purpose of this test is to test whether female and male have equal mean of final grading. The two sample contain 208 Female and 187 Male after being sorted out. The population variance is unknown and assumed to be unequal. The test is being conducted at α=0.05, where α is significant level of confidence.

Null Hypothesis: The mean quality of grades between female and male are equal.

Alternate Hypothesis: The mean quality of grades between female and male are not equal.

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Where,

μ_1 is mean quality of grades from female students and μ_2 is the mean quality of grades from male students

```
Gender
  F   M
208 187
> #Hypothesis testing between Gender and G3
> t.test(Grade3~Gender,alt="two.sided",mu=0,conf=0.95,var.eq=F,Paired=F)

        Welch Two Sample t-test

data:  Grade3 by Gender
t = -2.0651, df = 390.57, p-value = 0.03958
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.85073226 -0.04545244
sample estimates:
mean in group F mean in group M
       9.966346       10.914439

> qt(0.025,390)
[1] -1.966065
> -qt(0.025,390)
[1] 1.966065
```

Figure1: Hypothesis Testing of 2 sample mean using RStudio

Based on figure 1, Female mean based on grade 3=9.966346 while Male mean based on grade 3=10.914439. The degree of freedom =390.57 floor to 390. The test statistic,T_0 =-2.0651. The P-value of the test statistic is 0.03958

At α=0.05/2, critical value,$t_{0.025,390}$ =-1.966065 & 1.966065

Since T_0 =-2.0651<-1.966065 and P-value = 0.03958>0.025, we reject the H_0, null hypothesis. There is enough evidence to prove that the mean of Female grade is not equal to Male grade at significance level of 0.025.

2.Correlation

The variables that being used for Correlation test are study time and daily alcohol consumption. The purpose of this test the strength of relationship between the variables, using Spearman's Correlation Coefficient, the value of correlation coefficient is obtained.

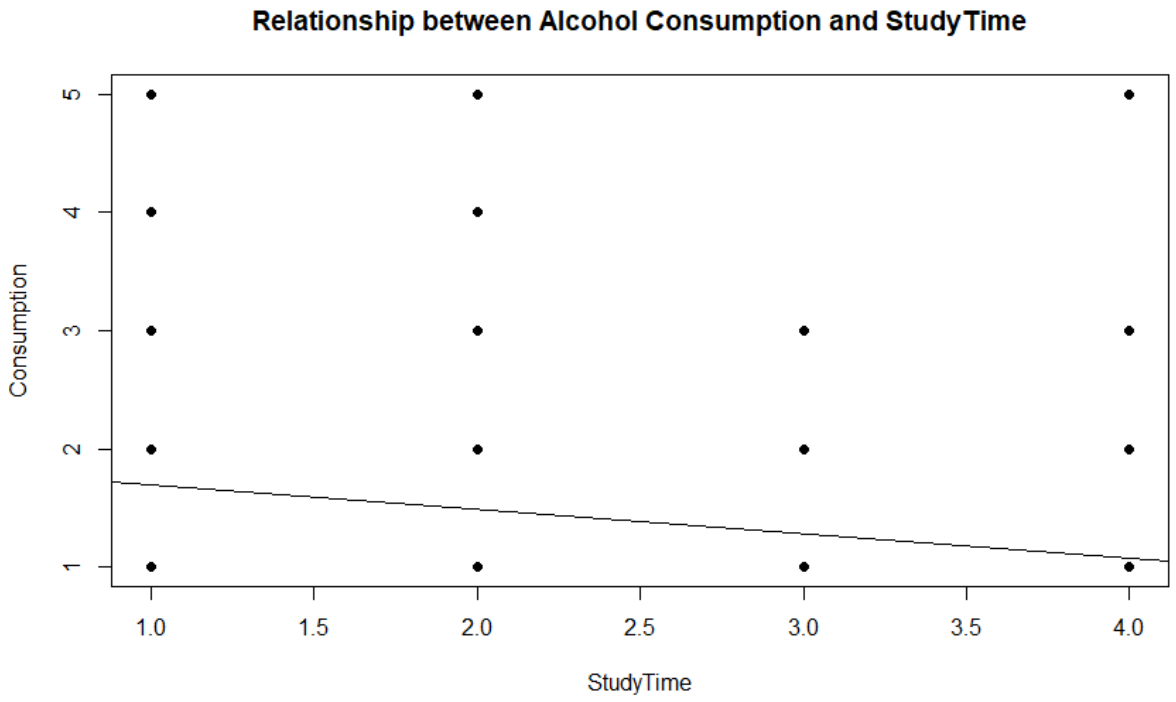

Figure 2: Relationship between Alcohol Consumption and Study Time

From the figure above, we can see that the Alcohol Consumption decreases as the Study Time decreases. The graph also proves that there is negative relationship between the variables of Alcohol Consumption and Study Time.

```
        Spearman's rank correlation rho

data:  Consumption and StudyTime
S = 12509793, p-value = 1.245e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.2179035

> plot(StudyTime,Consumption,pch=19,main='Relationship between Alcohol Consumption and StudyTime')
> abline(lm(Consumption~StudyTime))
> -qt(0.025,393)
[1] 1.966019
> qt(0.025,393)
[1] -1.966019
```

Figure 3: Spearman's Correlation Coefficient and test statistics being calculated in RStudio

Based on the figure 3, The correlation coefficient, r, is being calculated using the software called RStudio. The x-axis represents the Study Time while the y-axis represents the variable Alcohol consumption. Based on the figure above, r=-0.21790. The coefficient value shows that there exists a relationship between the variables but it is weak. To improve the accuracy of the test, we test the significant value for the correlation with α=0.05. The test is a two tailed test. Hence, α=0.05/2

Null Hypothesis: There is no linear correlation between Alcohol Consumption and Study Time

Alternate Hypothesis: There exist a liner correlation between Alcohol Consumption and Study Time.

$H_0: \rho = 0$

$H_1: \rho \neq 0$

Critical Value:$t_{(0.025,393)}$ =-1.966019 – 1.966019

Test Statistics, T=-0.2179035

P-value=1.24e-05

Since T=0.2179035<1.966019 & P-value=1.24e-05<0.025 r=0.21790, we reject the null hypothesis. There is enough evidence to prove that there exists a linear correlation between Alcohol Consumption and Study Time at significance level of 0.05.

3.Regression

In this 3rd test, we are going to test the relationship between Absence and Grade 3. The independent variable is Grade 3 while dependant variable is Absence.

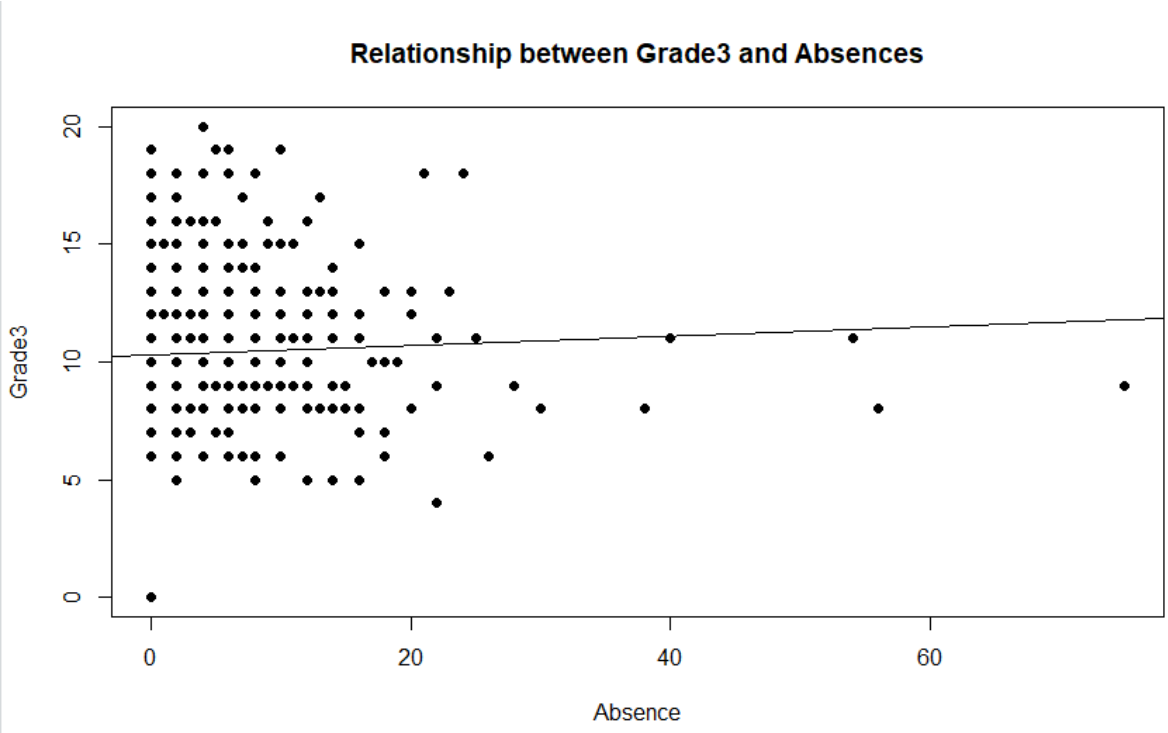

Figure 4: Relationship between Grade3 and Absences

Based on Figure 4, it can be concluded that the relationship between Grade3 and Absences are appeared to be linearly positive. To increase the accuracy, linear regression model is used.

```
Call:
lm(formula = Grade3 ~ Absence)

Residuals:
     Min       1Q   Median       3Q      Max
-10.3033  -2.3033   0.5007   3.4811   9.6183

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.30327    0.28347  36.347   <2e-16 ***
Absence      0.01961    0.02886   0.679    0.497
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.585 on 393 degrees of freedom
Multiple R-squared:  0.001173,  Adjusted R-squared:  -0.001369
F-statistic: 0.4615 on 1 and 393 DF,  p-value: 0.4973

> -qt(0.025,393)
[1] 1.966019
> qt(0.025,393)
[1] -1.966019
```

Figure 5: Summary of linear regression model based on RStudio

Based on Figure 5, we obtain the estimated regression models as below,

$$\hat{y} = 10.30327 + 0.01961x$$

Where,

$\hat{y}$ = Grade3

x = Absences

$b_0 = 10.30327$

$b_1 = 0.01961$

Based on the equation above, we can conclude that the average value of y=10.30327 at x=0. As x increase by 1, we obtained a change of average value of y=0. 01961.Other than that, coefficient of determination,R^2 is also calculated. Since R^2=0.001173,0<R^2<1, There exist a weak linear relationship between variable x and y. To test the regression, at level of confidence, α=0.05 is being used. The test is two tailed tests. Hence, α/2=0.025

Null Hypothesis: There is no linear relationship between Grade3 and Absence

Alternate Hypothesis: There exist linear relationship between Grade 3 and Absence

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Critical value: $t_{(0.025,393)}$ =-1.966019 – 1.966019

Test statistic, t=0.679

P-value= 0.4973

Since t=0.679<1.966019 & P-value=0.4973>0.05, We failed to reject the null hypothesis. We have insufficient evidence to prove that there is no linear relationship between the variables Grade 3 and Absence at significance level of 0.05.

4.Chi-Square test of Independence

The main purpose of Chi-Square test of independence used is to find out whether there is a relationship between two variables that is Gender and Higher Education which is a nominal data. The data consist of two types of answer for each variable which will produce df=1.

```
> #chi Square two contigency test between Gender and HigherEducation
> Chi=table(Table[,3],Table[,22])
> Chisquare<-chisq.test(Chi,correct=FALSE)
> alpha<-0.05
> Chi.alpha<-qchisq(alpha,df=1,lower.tail=FALSE)
> Chisquare$observed

     no yes
  F   4 204
  M  16 171
> Chisquare$expected

          no      yes
  F 10.531646 197.4684
  M  9.468354 177.5316
> Chi.alpha
[1] 3.841459
> Chisquare

        Pearson's Chi-squared test

data:  Chi
X-squared = 9.013, df = 1, p-value = 0.002681
```

Figure 6: Calculation of Chi-Square Test of Independence using RStudio

Based on Figure 6 above, The Chi-Square test of Independence is conducted at significance level of α=0.05.

Null Hypothesis, H_0: Variable Gender and Higher Education are independent

Alternate Hypothesis, H_1: Variable Gender and Higher Education are dependant

Critical Value,$X_{(0.05,1)}{}^2$ =3.841459

Test statistic,X^2 =9.013

P-value=0.002681

Since X^2 =9.013>3.841459 & P-value=0.002681<0.05, we reject H_0,null hypothesis. There is enough evidence to prove the Gender and Higher Education are dependant at significance level 0.05.

Discussion

Firstly, hypothesis testing of 2 sample mean, we can conclude that Male mean of Grade 3 is higher than Female mean of Grade 3. This is male are less vulnerable to mental state health problem compared to female. Mental state does affect the study performance of the students especially the female students. Other than that, factors like education of father and mother also affects the study grade of the students. Parents with a better background education have higher chances to obtained a better grade compared to other students with a poor background of studies.

Secondly, Correlation coefficient test, we can conclude that Alcohol consumption does have linear correlation with Study time. Although the correlation is weak, still there exist a relationship between the two variables. This is due to the fact that if Study time increases, the consumption of alcohol of alcohol increases because the free time is occupied with study. Hence, there is no free time for the students to spend their time to consume alcohol which might affects the study performance of the students. Spearman's Correlation Coefficient is being used because the data type of both variable is ordinal. Hence to obtain a better accuracy of correlation efficient, this particular method is being chosen.

Thirdly, based on regression test, we can conclude that there exists a weak relationship between the variable of Absences and Grade3. The higher the amount of absences, the higher the Grade 3 based on the graph. This is due to the fact that other factors such as students with a better family background and less family problems will somehow affect the student performance. Although number of absences is high, it will depend on other factors such as family relationship.

Lastly, based on the Chi-Square test of independence, we can conclude that the Gender and Higher Education are dependant with each other. This is because Female tend to pursue more on higher education because female wants to pursue their career goals compared to man which prefer to work right after high school education.

Conclusion

In conclusion, we can conclude that alcohol does not affects the student's performance that much because there still exist students who consumed alcohol heavily but still be able to get a good result for their final grade. Hence, the performance of the students is affected by other factor such as family background and mental health of the particular students. Lastly, alcohol does bring positive and negative impacts, but in this particular case study, there is no enough evidence to prove that alcohol does affects the study performance of students as there is still a lot of students who does not consumed alcohol but still doesn't get a good grades for their final exam.

References

Boston University School of Public Health. (6 January, 2016). *Correlation*. Retrieved from spchweb: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R5_Correlation-Regression/R5_Correlation-Regression3.html

Kassambara, A. (n.d.). *Correlation test between two variables in R*. Retrieved from Stastical tool for high throuhput data analysis: http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r

Ortiz, J. (23 September, 2016). *Student Alcohol Consumption*. Retrieved from Data.World: https://data.world/databeats/student-alcohol-consumption