



Students Performance in Mathematics Subject

Name:Ragu Raju Naidu A/L S Palani
Raju Naidu

Matrik: A19EC0152

Course:PSDA/ SECI2143

Section: 03

Abstract

The objective of this study is to investigate the relationship of average performance in mathematics of Gabriel Pereira secondary school students by performing and finding inferential statistics such as hypothesis testing , correlation and regression model between these two variables .

I Introduction

Mathematics is on of the subjects that is offered to students from all around the world. Students will learn many elements of mathematics such as arithmetic , computation , algebra , geometry , calculus , probability and logic . Mathematics helps students to learn alot of things that we used in daily life. For example, it helps us to train our brain to perform better at multitasking , able us to read clock, help us to manage our personal

finance, help us to solve problems effectively, help us to discover real world formulas , theories that push our civilization for further growth .

Mathematics subject is hence, designed under curricular system to provide students with necessary mathematical skills that they can apply in real life.

As the students progress from primary school classes to secondary school, students often find it that mathematics subject getting trickier , and difficult to understand. Most of the students struggle alot in mathematics subjects especially with the introduction of algebraic equation in upper secondary school. Fur the more , many

students lost chances to good ranking just because, some of the students can't perform we in mathematics, while some of the students perform well in Mathematics subject naturally.

This study aims to investigate the average performance of the students in Mathematics subjects and identify the causes and affect their performance in Mathematics subject.

II - Methodology

The data in the study is obtained via an open source website, Kaggle which provide it's user with many data and analysis on various matters. The data were obtained from a survey of student's performance in Mathematics and Portuguese subjects. The data contain alot of informations such as , gender, father's and mother's education, family members, studying hours.

There are 349 data is in total from the dataset . Hypothesis testing was done to determine if there's sufficient statistical proof to support the null hypothesis. The sample is normally distributed and plotted using R Studio.

III –Results&Discussion

A.Hypothesis testing - Based on the samr data obtained from the dataset " Students Performance in Mathematics Subject" , we claim that the average performance of students in Mathematics subject in G1 is not equal to the average average performance of students in Mathematics subject in G2. Where u_1 is the performance in G1 and u_2 is the performance in G2. n_1 is the number of data Performance in G1 while n_2 is the number of data Performance in G2. A sample of 349 data were chosen to calculate the P-value at 0.05 level of

significance. The formula as shown below we're used for t-test

$$H_0: u_1 = u_2$$

$$H_1: u_1 \neq u_2$$

The test statistic, t is used and data sample, $n = 349$. We obtain t equals to 0.7723 with the value $u_1 = 10.90886$ and $u_2 = 10.71392$. The critical value was found using the t-table with degree of freedom, $df = 775.98$.

Below is the result in r where $x\$G1$ is the Performances in G1 and $x\$G2$ is the the Performances in G2 ,

```
> t.test(x$G1,x$G2)
```

Welch Two Sample t-test

data: x\$G1 and x\$G2

$t = 0.7723$, $df = 775.98$, $p\text{-value} = 0.4402$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.3005526 0.6904260

sample estimates:

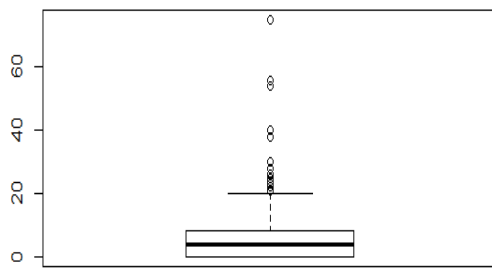
mean of x mean of y

10.90886 10.71392

The test statistics, $t <$ than critical value , at 0.05 significance level . Thus we failed to reject the null hypothesis. There's sufficient evidence to claim that the mean of the performance of students in Mathematics subject for G1 is equal to G2 at a 95% confidence level

B. Correlation

Based on the correlation test, we can analyse the relationship between the performance of students in Mathematics subject for G1 test and the absence and how intensely the number of absence affects students performance. 349 data from the dataset we're selected to find the outlier with the help of boxplot.



With the analysis made from the boxplot, we can remove 16 data since being outliers in the data. Thus, we have a total numbers of N=333, to identify the correlation coefficient. The data type is ratio and ordinal. Thus, Pearson's technique were used to find the correlation coefficient, r . The variable x is, the students performance in Mathematics Subject for G1 while, the variable y is the number of

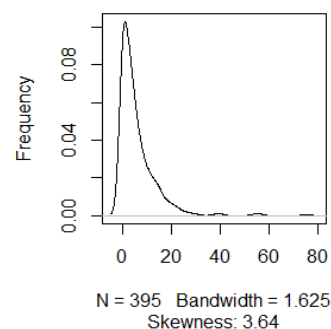
absence. The formula used was:

Correlation Coefficient Formula

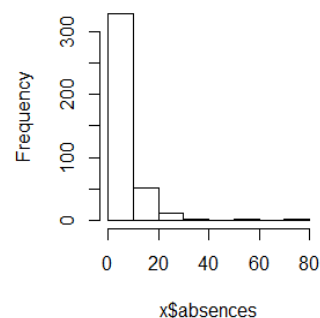
$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

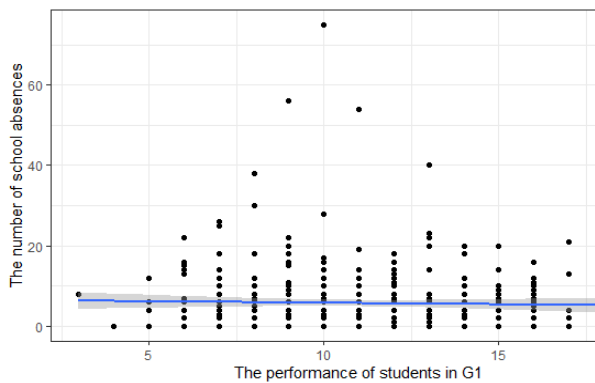
The correlation coefficient, r is 0.01967005. With the results, we can understand that the students performance in mathematics subjects for G1 has a weak positive correlation with number of absence. This means, the higher the absences the lower the performance in G1. This can be seen from the scatter plot, where the dots form a weak positive slop.

Density Plot: Absences



Histogram of x\$Absences





Pearson's product-moment correlation

```
data: log(x$absences + 1) and log(x$G1 + 1)
t = 0.39002, df = 393, p-value = 0.6967
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.0791546  0.1181118
sample estimates:
cor
0.01967005
```

B.1. Significance Test for Correlation:

The 2 variables are then tested to find any evidence of linear between the school absences and the performance of students in Mathematics subject for G1. The significance level is at 0.05.

: The significance test as below:

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

$H_1: \rho \neq 0$ (linear correlation exist)

: The formula used is :

$$t = \frac{r\sqrt{n}}{\sqrt{1-r^2}}$$

The sample size is, $n = 333$, $r = 0.01967005$. From the significance test made in R, we found that the value of the t is equal to 0.39002.

The sample size is, $n = 333$, $r = 0.01967005$. From the significance test made in R, we found that the value of the t is equal to 0.39002.

Since the test statistic, $t <$ critical value we failed to reject the null hypothesis at $\alpha = 0.05$. There is sufficient evidence to claim that, there's a linear relationship between school absences and performance of Students in Mathematics for G1. Through this test, we cannot come to a conclusion that, the number of absences will have a direct affect on students performance in mathematics Subject for G1. This is because, there might be some other reasons that affects students performance in Mathematics. Moreover, the number of absences might actually, resulted from health issues, thus they could actually have studied at their health care centres. Finally, the number absences itself could have included the numbers a student went for Extra curricular activities or other academic competitions.

C. Linear Regression.

The regression of the two variables are analysed to predict a mathematical equation between the school absences and the Performances of students in Mathematics Subject for G1 Test. 333 data from the dataset used again for this test. The variable x is , the performances of students in Mathematics Subject for G1 Test while y is the number of absences. The Formula as below:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

y is dependent variable,

β_0 is the population intercept y,

β_1 is the population slope coefficient

x is independent variable,

ε is random error component.

Call:

```
lm(formula = x$absences ~ x$G1, data = x)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.225	-5.403	-2.001	2.186	69.223

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.52433	1.38605	4.707	3.49e-06 ***
x\$G1	-0.07475	0.12157	-0.615	0.539

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.009 on 393 degrees of freedom

Multiple R-squared: 0.0009612, Adjusted R-squared: -0.001581

F-statistic: 0.3781 on 1 and 393 DF, p-value: 0.539

The regression model was calculated as shown above. We obtain $Y = 6.5243 - 0.07475x$. Moreover,

the coefficient of determination, R^2 is calculated to see the portion of the total variation in the dependent variable, the school absences is explained by the variation in the Performances of students in the Mathematics subject for G1 Test. The R-square obtained was -0.001581 which is 6% of the variation.

Thus, this shows us that , the number of absence has little effect on the performances of Students in Mathematics Subject for G1 Test. This is because, students who takes absences might have took it for academic competitions such Maths Olimpiad, where they actually spend more time to answer Maths questions. Moreover, health factors play

role in number of absences, but it doesn't mean that students with health problems can't perform well in Maths test in G1.

D. Chi Square Test

A Chi-Square test was performed to test how likely it is that the observed distribution is due to chance where t1 is the table of Performances in Mathematics subject for G1 test and family. We can also use this test to determine whether the observed distributions of data could fit the expected distributions. The null hypothesis, H0 and alternative hypothesis, H1 as below:

Ho: There is no relationship between family size and the performances of students in Mathematics Subject for G1 Test and both variables are independent to one another.

H1: There is a relationship between family size and the Performances in Mathematics Subject for G1 test and both variables are dependent of each other.

The following formula was used to obtain the test statistics for Chi-Square test.

The expected count Eij is calculated using this formula:

$$E_{ij} = \frac{(i^{\text{th}} \text{ Row Total})(j^{\text{th}} \text{ Column Total})}{\text{Total sample size}}$$

The observed value is as shown in figure 6, the test statistic x2 is calculated using R Studio.

$$\text{Test statistic, } x^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

```
Pearson's Chi-squared test

data: t1
X-squared = 15.003, df = 16, p-value = 0.5244
```

```
> view(t)
> chisq.test(t)$expected
```

	GT3	LE3
3	0.7113924	0.2886076
4	0.7113924	0.2886076
5	4.9797468	2.0202532
6	17.0734177	6.9265823
7	26.3215190	10.6784810
8	29.1670886	11.8329114
9	22.0531646	8.9468354
10	36.2810127	14.7189873
11	27.7443038	11.2556962
12	24.8987342	10.1012658
13	23.4759494	9.5240506
14	21.3417722	8.6582278
15	17.0734177	6.9265823
16	15.6506329	6.3493671
17	5.6911392	2.3088608
18	5.6911392	2.3088608
19	2.1341772	0.8658228

From the table above, GT3= student who have family member greater than 3. LE3=student who have family member lesser than 3. The value of X squared is 15.003 and the value of degree of Freedom is 16. The critical values are found to be approximately 26.296 from the Chi-Square Distribution Table. Since the test statistic, x-squared < critical value (15.003 < 26.296), the test statistic, x-squared didn't

fall within the critical region, therefore we fail reject the null hypothesis. There is sufficient evidence to show that the family size is independent to the Performances in Mathematics Subject in G1 test.

Thus, the results obtained, shows that the family size also have no effect on Student's performances in Mathematics Subject for G1 Test. The performances of students might actually have related to their Individual effort and their habits, friends circle and their Parents education Status. Regardless, of the background of a particular student, he should have put more effort to perform well in Mathematics Subject since Mathematics is all about doing Practices.

IV -Conclusion

From the hypothesis , The test statistics, $t < \text{than critical value}$, at 0.05 significance level . Thus we failed to reject the null hypothesis. There's sufficient evidence to claim that the mean of the performance of students in Mathematics subject for G1 is equal to G2 at a 95% confidence level.

From the correlation test, he correlation coefficient, r is 0.01967005. With the results, we can understand that the students performance in mathematics subjects for G1 has a weak positive correlation with number of absence. This means, the higher the absences the lower the performance in

G1. This can be seen from the scatter plot, where the dots form a weak positive slop.

From the regression test, he regression model was calculated as shown above. We obtaain $Y = 6.5243 - 0.07475x$.

From the Chi-Square Distribution, The value of X squared is 15.003 and the value of degree of Freedom is 16. The critical values are found to be approximately 26.296 from the Chi-Square Distribution Table. Since the test statistic, $x\text{-squared} < \text{critical value}$ ($15.003 < 26.296$), the test statistic, $x\text{-squared}$ didn't fall within the critical region, therefore we fail reject the null hypothesis. There is sufficient evidence to show that the family size is independent to the Performances in Mathematics Subject in G1 test.

V- Acknowledgement

A sincere thanks to the lecturer, Dr Aryati to helped me with my doubts.

VII –References

1. I.G. Betteley &R.J. Clarke 09 Jul 2006 The Academic Performance of Students on a Mathematics Degree Course [International Journal of Mathematical Education in Science and Technology](#)
2. Charles Y. C. Yeh 11 March 2019 Enhancing achievement and interest in mathematics learning through Math-Island <https://telrp.springeropen.com/articles/10.1186/s41039-019-0100-9>
3. dr. evaristo andreas mtitu factors leading to poor performance in mathematics subject in kibaha secondary schools <https://core.ac.uk/reader/44684738>