**Group Member 9:**

**Asmadi, Norshahirah**

**Rukin, Indah Puspita Dewi**

**Tarmizi, Mohammad Hafiz**

**Topic :**

Does caffeine affect academic performances among college students? Caffeine stimulates central nervous system of the body which makes a person more energetic and less tired for longer period. Due to constraint hours of study, most students chose to drink coffee as an alternative to stay awake in the night time. However, lack of sleep could lead to stress and drowsiness. Therefore, we would like to investigate the relationship between **caffeine, amount of sleep, and academic performances in class among students.**

**Description of Data Sets:**

20 students of Selesian University of Technology are randomly selected equally in a class to participate in the experiment. They were asked the average amount of coffee they consumed per night (per cup), the amount of their sleep (per hour), and their Grade in the same class. This data is collected for the whole week before an exam. The objective of this observation survey is to prove one of these statements:   
1. More coffee lead to better academic performance  
2. More coffee lead student to minimize amount of sleep (as a substitution to sleep)

3.         More sleeping hours lead to a better grade.

The original dataset can be obtained from:

<https://www.statcrunch.com/5.0/viewreport.php?reportid=36611#Data1>

Question to be addressed (survey):  
1) What is your Grade in the class?  
2) On average, how many cups of coffee do you drink in one day?  
3) On average, how many hours do you sleep at night?

**DATASETS**  
**Number  Cups        Sleep Exam grade**1      5 40      70

2      8 34       80  
3      9 29       85

4     11 31       54  
5     20 30       95  
6     10 24      100  
7      9 27       89  
8      7 28       67  
9      9 35       98  
10      7 33       77   
11     10 38       68  
12      7 34       76  
13     12 18       49  
14       1 39       90  
15       9 33       78  
16      12 20       38  
17       5 28       78  
18     11 22      41  
19      8 32      59  
20     10 19      60

**SAS code and result:**

**data** coffee;

input cup sleep grade;

datalines;

5 40 70

8 34 80

9 29 85

11 31 54

20 30 95

10 24 100

9 27 89

7 28 67

9 35 98

7 33 77

10 38 68

7 34 76

12 18 49

1 39 90

9 33 78

12 20 38

5 28 78

11 22 41

8 32 59

10 19 60

;

**run**;

We want to check correlation of the variables by using *proc corr* in SAS, and we will do one-sided hypothesis testing (based on the correlation) for checking the linear relationship for these variables by using *proc reg* at significance level of 0.05.

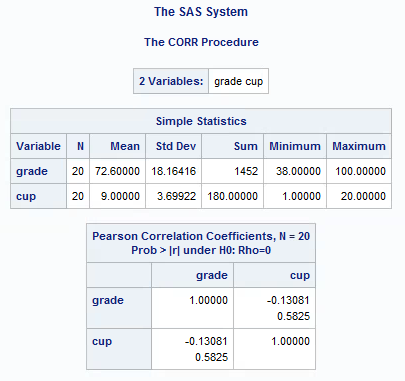
1. **Correlation for number cup of coffee and grade**

SAS Coding:

**proc** **corr** data = coffee;

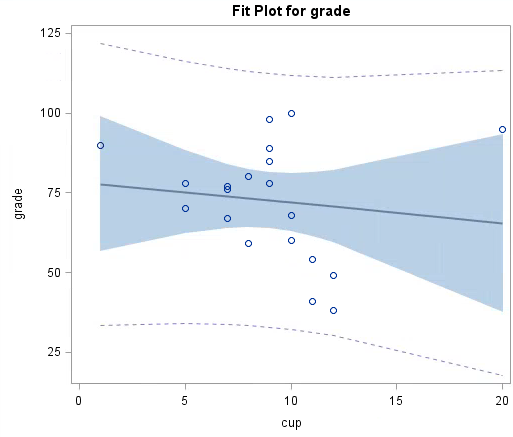
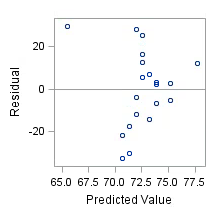
var grade cup;

**run**;



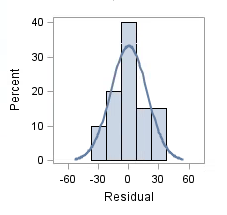
Checking the conditions of inference:

1. Linear relationship.

Based on the figures above, there is roughly negative linear relationship between these variables.

1. Normal variation about the line



As we can see, the histogram of the residuals is normal.

1. Independent

Based on the scatter plot and residual plots, we can see that all 20 observations are independent since there is no overlapping values.

1. Spread about the line stays the same

As we can see in the residual plots, the spread among the observations is about the same.

Since the condition is met by the data, we can proceed with our hypothesis testing on the slope of the linear regression line.

Hypothesis testing statement:

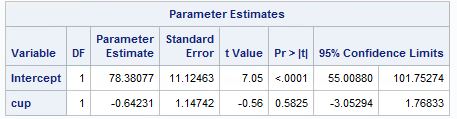
H0: =0

Ha: <0

**proc** **reg** data = coffee;

model grade = cup/clb;

**run**;



P-value = 0.58252/2= 0.29125

Claim: More coffee lead to better academic performance

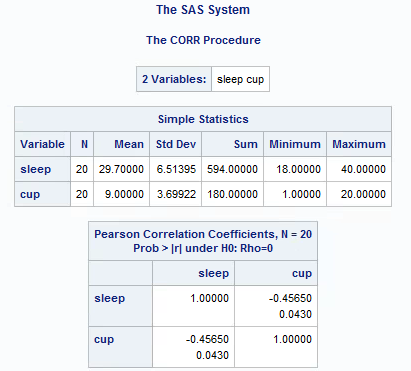
From the hypothesis test, the p-value is 0.29125. We fail to reject the hypothesis since 0.29125 > 0.05. Additionally, as we can see the 95% confidence interval of parameter estimate or the slope, (-3.05294, 1.76833), contained zero. This is another reason we fail to reject the null hypothesis. Hence, there is no evidence to conclude there is negative association between the number cup of coffee and the academic performance, or we could say that there is no correlation between these quantity variables. Because of this, we could not prove statement (i).

1. **Correlation for number cup of coffee and duration of sleep**

**proc** **corr** data = coffee;

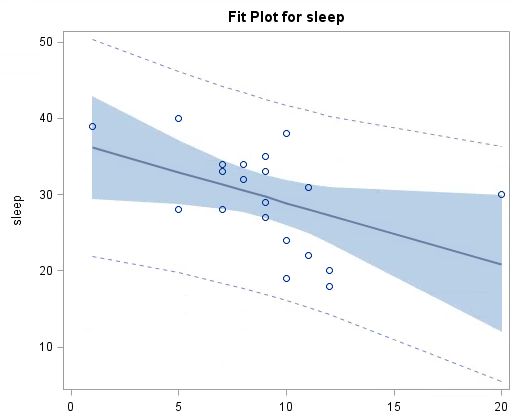
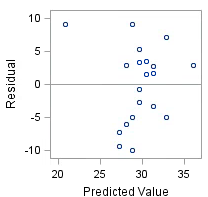
var sleep cup;

**run**;



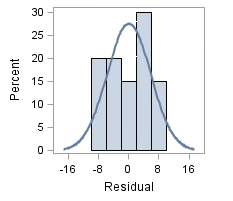
Checking the conditions of inference:

1. Linear relationship.

Based on the figures above, there is roughly negative linear relationship between these variables.

1. Normal variation about the line



As we can see, the histogram of the residuals is roughly normal.

1. Independent

Based on the scatter plot and residual plots, we can see that all 20 observations are independent since there is no overlapping values.

1. Spread about the line stays the same

As we can see in the residual plots, the spread among the observations is about the same.

Since the condition is met by the data, we can proceed with our hypothesis testing on the slope of the linear regression line.

Hypothesis testing statement:

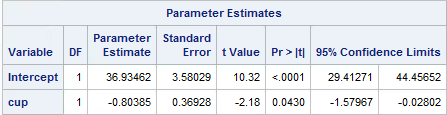
H0: =0

Ha:  <0

**proc** **reg** data = coffee;

model sleep=cup/clb;

**run**;



P-value = 0.0432= 0.0215

Claim: More coffee lead student to have less sleep

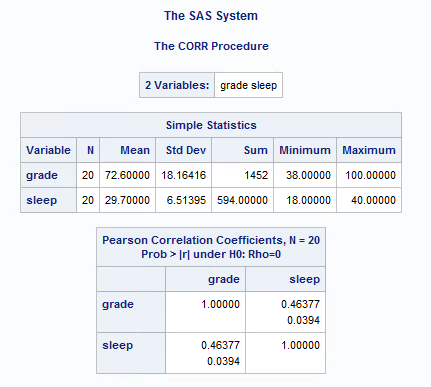
According to SAS result, the p-value for this data is 0.0215. Since the p-value is lower than significant value of 0.05, we reject the null hypothesis. Another reason we reject the null hypothesis is the 95% confidence interval for the slope, (-1.57967,-0.02802), does not contain zero in the interval. There is significantly evidence that more coffee leads student to have less sleep hour. Furthermore, the correlation value for these variables is -0.4565. We can conclude that there is negatively moderate correlation as the correlation is near to 0.5. Based on the correlation, the statement (ii) might be true, but since it is correlation we cannot say much. The reason for this is there might be some lurking variables which can support the statement (ii). Some of the lurking variables:

1. Entertainment as substitute for sleep.
2. Hard to sleep because of depression
3. Insomnia
4. **Correlation for duration of sleep and academic performance**

**proc** **corr** data = coffee;

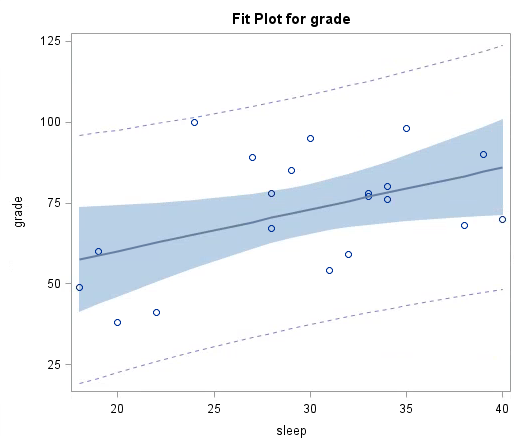
var grade sleep;

**run**;



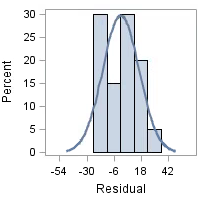
Checking the conditions of inference:

1. Linear relationship.

Based on the figures above, there is roughly positive linear relationship between these variables.

1. Normal variation about the line



As we can see, the histogram of the residuals is roughly normal.

1. Independent

Based on the scatter plot and residual plots, we can see that all 20 observations are independent since there is no overlapping values.

1. Spread about the line stays the same

As we can see in the residual plots, the spread among the observations is about the same by excluding the outliers.

Since the condition is met by the data, we can proceed with our hypothesis testing on the slope of the linear regression line.

Hypothesis testing statement:

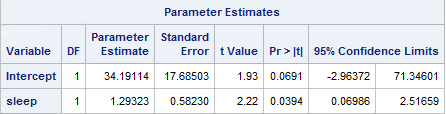
H0: =0

Ha: >0

**proc** **reg** data = coffee;

model grade=sleep/clb;

**run**;



P-value = 0.03942= 0.0197

Claim: More sleep hours boost academic performance

From the hypothesis test, the p-value is 0.0197. At =0.05, we will reject the null hypothesis because the p-value = 0.0197 < 0.05. Moreover, the 95% confidence interval of the parameter of interest in the hypothesis, (0.06986, 2.51659), does not contain zero in between the interval, so it is another reason we reject the null hypothesis. Therefore, there is strong evidence to against the null hypothesis. Also, we can conclude there is a positive linear relationship between the sleep hours and academic performance. Based on the SAS result above, the correlation between sleep hours and academic performance is 0.46377. Hence, these variables have moderately positive correlation. Since correlation only proves association between the quantity variables, we cannot conclude statement (iii). There might be some lurking variables that confounded the relationship of variables in statement (iii). Below are listed some of the lurking variables:

1. Understanding of the course materials or homework given
2. Students’ interest in the course
3. How often the students go to tutor session?

**Conclusion**

From this sample of data, we do not able to conclude any of the statement we stated above. However, we can prove the association two out of three association pair of variables. Based on this data, we can conclude these statements:

1. There is no correlation between the number of cups of coffee taken and the grade achieved in the exam.
2. There is positive association between the hours of sleep and the grade received by the students.
3. There is negative association between the number of cups taken and the sleep hours of the student.

**Remaining Tasks:**

We have agreed to present our report by posting the document on the course web page.

**Itemized lists of specific tasks done by each member:**

Indah Puspita Dewi - Description of the observation survey, code computing

Mohd Hafiz bin Tarmizi - Came up with methods applied, interpretation of SAS output

Norshahirah Asmadi - Found the datasets to be used, editing the report.