MSCI 718 – INDIVIDUAL ASSIGNMENT 2

Data Summary:

This dataset gives information about medical students' mental health such as burnout, anxiety, empathy and depression in Switzerland. This dataset has 20 variables with 886 objects. This data set has both categorical and continuous variables. However in this report we find out if there is any relationship between the **age of the student**(age; min: 17, max: 49, mean:22.38, median: 22) with **Hours of study per week of the participant**(stud_h; min: 0 , max:70, mean:25.29, median: 25), **Maslach Burnout Inventory-Exhaustion scale of the participant**(mbi_ex; min: 5, max: 30, mean: 16.88, median:17).

There is no data cleaning required in the dataset of our analysis as there are no missing values or outliers.

Planning:

Our analysis is to find the relationship between **age of the student** vs **Hours of study per week**, **age of the student** vs and **Maslach Burnout Inventory-Exhaustion scale**. In order to check the normality of the dataset, I plotted both QQ plots and histogram and also ran Shapiro-Wilk's test for all the variables and it was found that the variable **age** is highly positively skewed. Therefore that variable is transformed using log transformation.

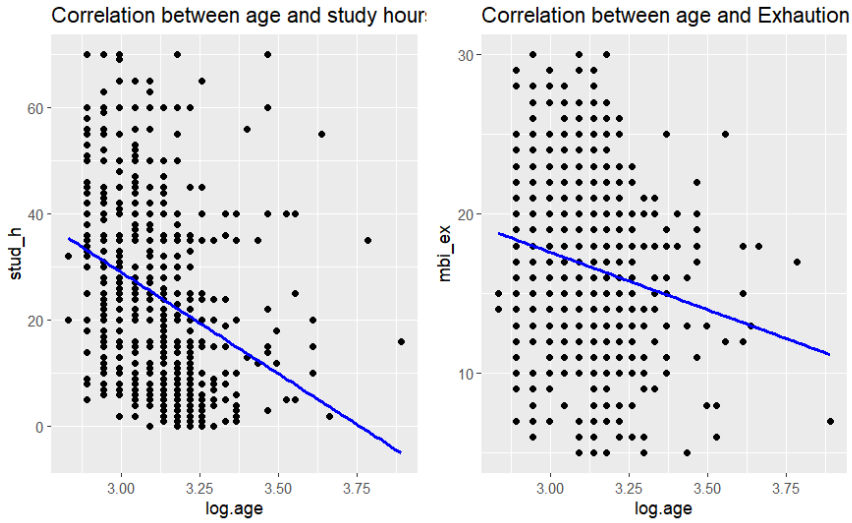
Even after transformation, the variable is not normalised. However our dataset has more than 30 datapoints, according to Central Limit Theorem the mean of our dataset is normalised if the dataset contains more than 30 datapoints. Therefore we proceed with Pearson's correlation.

Analysis:

Pearsons' Correlation between **log of age of the student vs Hours of study per week** is tested and it is seen that Hours of study is negatively significantly correlated with log of age of the student, r = -0.32 *(*p < 0.001, 95% CI -0.3807000 to -0.2626587). A correlation of -0.32 represents a medium-to-large effect explaining 10.24% of the variance.

Pearson's Correlation between **log of age of the student vs Maslach Burnout Inventory-Exhaustion scale** is tested and it is seen that Hours of study is negatively correlated with log of age of the student, r = -0.18 ($p < 0.001$, 95% CI -0.2475192 -0.1202674). A correlation of -0.18 represents a medium-to-large effect explaining 3.24% of the variance.

The below shown is the correlation graph.



Confidence intervals were also calculated using Bootstrap technique with type = "bca" to see if there is any significance difference from the one obtained from the result of Pearson's correlation. It can be seen that, there is no significant difference between the CI calculated in the two different ways.

The partial correlation between all the three variables were tested. The partial correlation between **log of age of the student, Hours of study per week and Maslach Burnout Inventory-Exhaustion scale** is -0.298878 and the variance that is shared is 8.93%.

There is a no significant relationship between log of age of the student and Hours of study per week ($r=-0.298878, p=1.064969e-19$) when controlling for the effect of Maslach Burnout Inventory-Exhaustion scale. This is slightly greater than the full correlation -0.32, which explained 10.24% variance.

<u>Conclusion:</u>

To find the correlation between age of student, hours of study per week and MBI exhaustion scale, I transformed "age" as it was not normally distributed. Then performed Pearsons' correlation on the variables and found that age has medium negative effect on hours of study per week and a very low negative effect on MBI exhaustion scale. From the partial correlation done for all the three variables, we can conclude that age has a medium negative effect on study hours per week, but there is a complex relationship between age of student, hours of study per week and MBI exhaustion scale

APPENDIX:

Summary of variables of our analysis:

```
      age            stud_h          mbi_ex
 Min.   :17.00   Min.   : 0.00   Min.   : 5.00
 1st Qu.:20.00   1st Qu.:12.00   1st Qu.:13.00
 Median :22.00   Median :25.00   Median :17.00
 Mean   :22.38   Mean   :25.29   Mean   :16.88
 3rd Qu.:24.00   3rd Qu.:36.00   3rd Qu.:20.00
 Max.   :49.00   Max.   :70.00   Max.   :30.00
```

Pearson's Correlation result:

```
        Pearson's product-moment correlation

data:  log.age and students_data$stud_h
t = -10.145, df = 884, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3807000 -0.2626587
sample estimates:
       cor
-0.3229347


        Pearson's product-moment correlation

data:  log.age and students_data$mbi_ex
t = -5.5866, df = 884, p-value = 3.083e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2475192 -0.1202674
sample estimates:
       cor
-0.1846672


        Pearson's product-moment correlation

data:  log.age and students_data$stai_t
t = -2.7698, df = 884, p-value = 0.005727
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.15765513 -0.02705846
sample estimates:
        cor
-0.09275571
```

Bootstrap CI calculation result:

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 2000 bootstrap replicates

CALL :
boot.ci(boot.out = bootPearsonResult, conf = 0.95, type = "bca")

Intervals :
Level       BCa
95%   (-0.3833, -0.2522 )
Calculations and Intervals on Original Scale
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 2000 bootstrap replicates

CALL :
boot.ci(boot.out = bootPearsonResult, conf = 0.95, type = "bca")

Intervals :
Level       BCa
95%   (-0.2463, -0.1225 )
Calculations and Intervals on Original Scale
```