Question: Is there any relation between the cognitive empathy and academic motivation among males and does it rely upon whether the person has a paid job or not?

Data Summary:

The Medical Student Mental Health dataset [1] explores medical students' empathy, mental health, and burnout in Switzerland. This dataset contains around 886 rows of data spread across 20 columns. It provides information about demographic factors such as age, sex and language spoken as well as internal measures such as job satisfaction, psychological distress, education grades, self-reported health status and empathy rating scales.

Out of these 20 parameters, some notable ones that I've used for my analysis are: -
- QCAE_COG: Cognitive Empathy score. Numeric datatype and continuous in nature.
- AMSP: Academic Motivation score. Numeric datatype and continuous in nature.
- JOB: Having a paid job? [0=No; 1=Yes] Numeric datatype and dichotomous in nature.

Data Cleaning:

I read the dataset and analysed it by plotting scatterplots and summarizing variables. I checked the dataset for null values, but I didn't find any. Then as per my problem statement, I filtered out the dataset to choose only Male gender data. Additionally, I selected the required variables: 'amsp', 'qcae_cog' and 'job' out of the whole dataset. Also, I analysed the scatter plots for these 3 variables to find any outliers, but there weren't any.

Choice of Variables for Correlation:

I used 'corrplot' command to help visualize a correlation matrix for the whole dataset. This helped me choose Academic Motivation Score (AMSP) and Cognitive Empathy Score (QCAE) as the target variables for my correlation analysis. From their respective scatter plots, both variables displayed continuous behaviour, and on testing their normality, further, I would be able to use them for Pearson's Correlation.

Checking if our data is Normally distributed:

We have assumed normality of our 2 target variables as it'll be necessary for our Pearson Correlation test.

I looked at 2 different graphs to see if our data was normal: Histogram and Q-Q plot. The histogram curves for both the variables appeared to be normal. The Q-Q plots also displayed near normality with slight deviation near the endpoints. As can be seen in Figure 1, on the right.

To confirm normality further, I performed some tests. For the Shapiro wilk test: p-value for AMSP = 0.1037 and for QCAE = 0.06102. Thus, both values denote significance. But we know this test is quite powerful and with large samples, the test can be significant. Thus, we should rely more on histogram, Q-Q plots, skewness, and kurtosis.

Q-Q plot for Academic Motivation Score and Cognitive Empathy Score
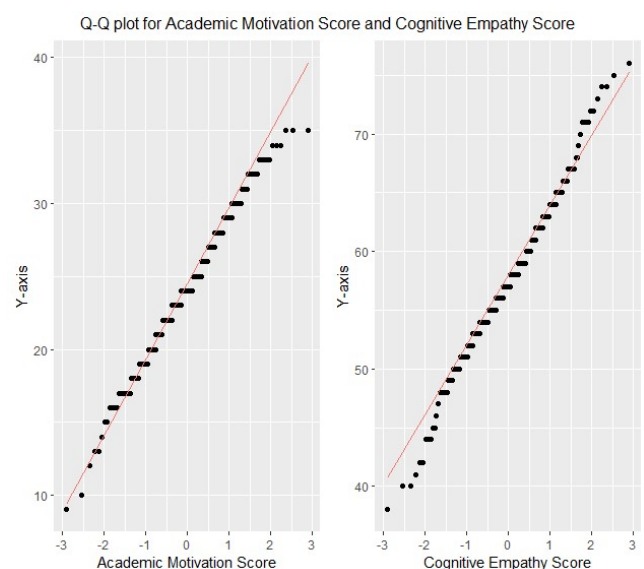
I used stat.desc() command under the pastecs library to evaluate skewness and kurtosis. Here, we utilize the normalized values "skew.2SE" and "kurt.2SE" obtained by dividing the skewness and kurtosis by 2 standard errors. These were: for AMSE: skew.2SE= -0.3860; kurt.2SE= -0.1636, and for QCAE: skew.2SE= 0.1822; kurt.2SE = 0.7712. We can see absolute values for both variables are smaller than 1. Thus, we can conclude with 95% confidence that skewness or kurtosis is equal to 0, and thus the distribution is normal. These values denote practical as well as statistical significance.

Planning and Analysis:

The scatter plot between Academic Motivation Score (AMSP) and Cognitive Empathy Score (QCAE_COG), Figure 2, appears to denote some positive correlation. Thus, I'll be using Pearson's correlation method to validate our assumptions: -

Relationship between Academic Motivation and Cognitive Empathy

- Data for both these target variables is normally distributed.
- A linear relationship exists between these 2 variables, and they are continuous in nature.

Hypothesis: **H0**: True correlation between AMSP and QCAE_COG is equal to 0; **H1**: True correlation between AMSP and QCAE_COG is not equal to 0.

Correlation: On running the Pearson's correlation test, we find that p-value < 2.2e-16. Thus, we reject null hypothesis in favour of the alternative that true correlation between AMSE and QCAE_COG is not equal to 0. Correlation score = 0.496188 denotes positive correlation between the variables. This means an increase in AMSE would also lead to increase in QCAE_COG. The closer this correlation value is to +1 or -1, the stronger is the correlation b/w those 2 variables. Thus, we have moderate positive correlation. Also, the obtained 95 percent confidence interval: [0.40, 0.58].

Partial Correlation: To test association between our target variables (AMSE and QCAE_COG) along with a controlling variable (job), we use pcor() and pcor.test() commands under the ggm library in R. We obtain pcor value= 0.4915507 with p-value = 4.49e-18. Thus, not much difference observed.

Conclusion: Shapiro wilk test found scores to be non-normal at 5% level of significance, however, inspection of Q-Q plots, skewness & kurtosis suggested that the data follows a normal distribution. Also, sample size was greater than 200, thus, we continued our analysis with the assumption of normality. From our analysis, we were able to conclude a moderately linear relationship between our target variables: Academic Motivation Score (AMSP) and Cognitive Empathy Score(QCAE_COG). Also, the obtained correlation value (0.496188) and partial correlation value (0.4915507) were similar which signified that there wasn't much effect of the controlling variable(job) towards the correlation between our target variables (AMSP & QCAE_COG).
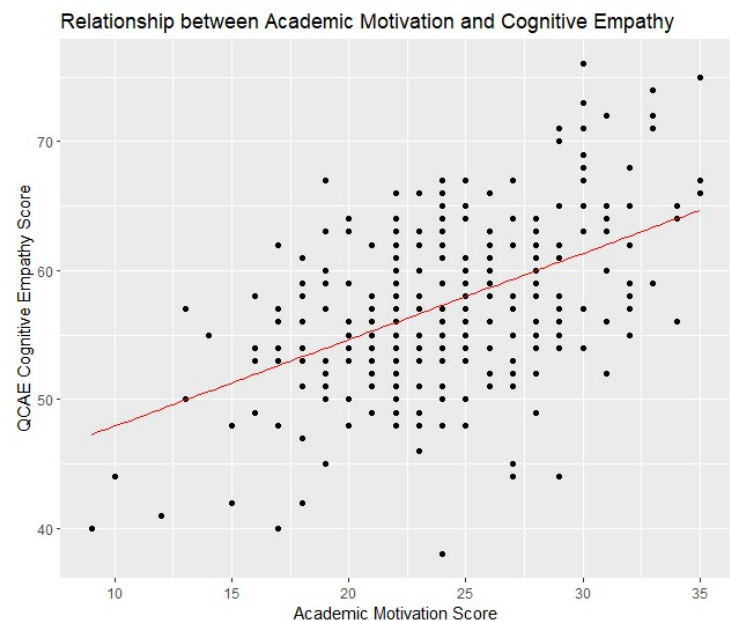
References:

[1] V. Carrard, "Dataset for the paper 'The relationship between medical students' empathy, mental health, and burnout: A cross-sectional study' published in Medical Teacher (2022)," *Zenodo*, Jul. 01, 2022. https://doi.org/10.5281/zenodo.5702895

[2] "Medical Student Mental Health," *Kaggle*, Jan. 25, 2023. https://www.kaggle.com/datasets/thedevastator/medical-student-mental-health

Appendix :

Correlation amongst all variables in our dataset: