Capstone I – Inferential Statistics

**Write a 1-2 page report on the steps and findings of your inferential statistical analysis. Upload this report to your GitHub and submit a link. Eventually, this report can be incorporated into your milestone report.**

Executive function (EF) is a cluster of cognitive processes that underlie planning, organizing, and regulation in humans, and is responsible for the achievement of purposeful, goal-directed behaviors. It affects the physical, mental, social, and psychological well-being of all individuals and is a very important factor in one's overall health. In the dataset, executive function is defined by three scores: interference, working memory, and cognitive flexibility.

**Problem:** Find major lifestyle trends in the data set related to executive function, identifying potential factors related to low executive function, and use these factors to build a predictive model.

First, the raw dataset was read into a Pandas DataFrame, and all missing values were dropped. An initial exploration of the dataset indicated that only 17 students were 14 years old (3.5%) and 33 were 16 years old (6.8%). The spread of 12, 13, and 15-year-olds was relatively even. Visual EDA followed and box plots were created for all categorical variables in the dataset which included sex, physical fitness category, physical activity category, global sleep category, income category, and household size. Scatter plots were made for the continuous variables (age, weight, BMI, BMI-for-age, physical fitness score, physical activity score, and sleep percent). The box plots were successful is showing the distribution of each category against each of the three EF scores, but the scatterplots were relatively inconclusive when it came to finding correlations between the variables and EF scores. A correlation matrix was graphed as well. However, it was necessary to perform statistical analysis to determine whether any perceived correlations were statistically significant.

Two-sample t-tests were conducted for two categorical variables - sex and global sleep category. The null hypothesis was rejected for sex with regard to cognitive flexibility (CF), signifying that the mean CF scores are statistically different for males and females. The null hypothesis was also rejected for global sleep category with working memory, suggesting that those who fall into the "good sleep quality" category exhibit statistically significant working memory scores than those who fall into the "poor sleep quality" category.

A different test was required for the remaining categorical variables, all of which had three or more categories within their respective variables. The One-Factor ANOVA was the most appropriate parametric test for the variables – however, the physical fitness and physical activity categories failed to meet the ANOVA condition regarding variances. As such, multiple non-parametric (Welch's) t-tests were conducted for both variables. The results were as follows:

- **PFS_CAT:** none of the categories have significantly different functioning scores than the means.
- **PA_CAT:** none of the categories have significantly different functioning scores than the means.

Multiple t-tests were conducted for income category with interference as well, and it was found that there is a significant difference between the means of Category 1 and Category 3 regarding interference.

The income category/working memory and income category/cognitive flexibility parameters met the conditions for ANOVA and the following results were obtained: 1) that there is a significant difference in mean working memory scores between students from low-income and high-income backgrounds, and 2) there is no significant difference between the three group means with respect to cognitive flexibility.

The Pearson correlation coefficient was a more apt test to determine the statistical significance of the continuous variables.

The following results were found:

- **Age:** There is a significant negative correlation between age and cognitive flexibility.
- **Weight:** There is no significant correlation between weight and executive function.
- **BMI:** There is no significant correlation between BMI and executive function.
- **BMI-for-Age:** There is a significant negative correlation between BMI-for-age and interference, as well as BMI-for-age and working memory. The correlation between BMI-for-age and cognitive flexibility is insignificant.
- **Physical_fitness_score:** There is no significant correlation between physical fitness scores and executive function.
- **PA_total_score:** There is no significant correlation between physical activity scores and executive function.
- **Sleep_percent:** There is a significant positive correlation between sleep percent and interference, as well as sleep percent and working memory.
- **Household_size:** There is a significant negative correlation between household size and working memory.