Predicting Executive Function Scores Using Linear Regression

I. The Problem

In the paper *Obesity and Unhealthy Lifestyle Associated with Poor Executive Function Among Malaysian Adolescents*, it was found that high BMI (body mass index) is associated with poor executive function.

Executive function is a cluster of cognitive processes that underlie planning, organizing, and regulation in the prefrontal cortex, and is responsible for the achievement of purposeful, goal-directed behaviors. In the literature, executive function is broken into three scores: interference (the ability to resist distraction and maintain focus), working memory (the ability to store, maintain, and manipulate information over a brief period of time), and cognitive flexibility (the ability to shift attention, select information, and alter response strategies in response to changing task demands) (Ying Hui Tee, et al. 1-2).

Proper executive function in childhood and adolescence is a very important factor in leading a healthy and balanced life through adulthood. Problems with functioning - executive dysfunction - can severely impact an adolescent's ability to learn and is a key marker in the development of Attention-Deficit/Hyperactivity Disorder.

My focus in this project is to determine which other lifestyle factors are predictive of low executive function scores in an effort to build a successful predictive model. The problem is as follows: finding major lifestyle trends in the dataset related to executive function, identifying potential factors related to low executive function, and using these factors to build a predictive model. The first step in the analysis of the dataset is to confirm whether the assumption regarding BMI scores is true, followed by an evaluation of lifestyle trends.

This data can be used by general practitioners in Malaysia to decide whether patients should be referred to specialists who can officially diagnose ADHD and prescribe medication. The state or local governments in Malaysia can also use the data to decide whether actions should be taken to regulate controllable factors such as physical activity levels and proper lunch options at school.

This study is a small one, but its effects hold implications on a global scale. While the initial experiment took place in Malaysia, if a similar study could be replicated in the United States for example, it could help determine what lifestyle modifications could be implemented nationally to improve executive function in students and tackle childhood obesity.

II. The Data

The dataset can be found in the Public Library of Science repository, a non-profit science and technology publisher with a library of open-access journals and literature. It is available as an Excel file with 513 subjects from two separate high schools in Selangor, Malaysia.

Any adolescents suffering from neurological or psychiatric disorders, medical conditions, learning disabilities, history of traumatic brain injuries, or who had difficulties in performing physical activities were excluded from the study. The students – aged 12 through 16 – provided answers to prompts regarding age, date of birth, sex, ethnicity, meal consumption patterns, physical activity, and sleep quality through self-administered questionnaires. The subjects' parents were given surveys to answer demographic information regarding household size, income category, father's years of education, and mother's years of education. Height and weight measurements were taken twice for each subject on-site at the time of experiment, with the mean value of each measurement used as the final measurement.

Three separate tests were conducted to measure executive function:

1. **Interference - (Stroop Color and Word Test)**

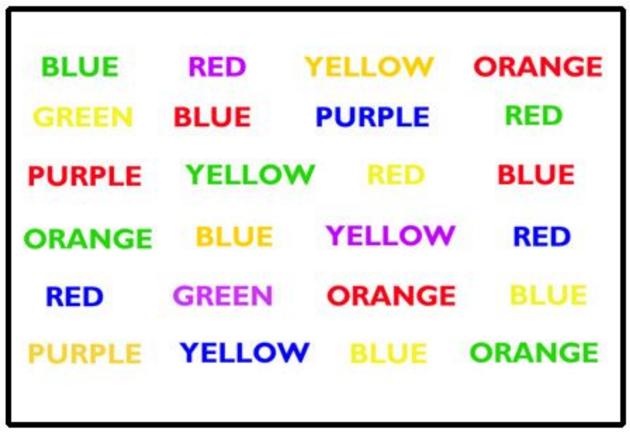


Figure 1. Example of the Stroop Color and Word Test

   a. A cognitive test based on the belief that individuals can recognize words faster than they can identify and name colors. Subjects are presented with a list of color words and must successfully name the ink color of the word presented, instead of the actual word, within a given amount of time (Fig. 1).
   b. A higher score denotes better capacity for inhibition responses, i.e. the ability to resist distraction and maintain focus.

2. **Working Memory - (Digit Span Test)**
   a. A test which measures short-term memory, in which the subject is given a sequence of numbers to memorize and repeat back. The sequence of numbers increases by one with each successive trial.
   b. Task performance is gauged by the average number of digits recounted correctly. A higher score denotes better working memory.

3. **Cognitive Flexibility - (Trail-Making Test)**
   a. A test in which subjects are given two pieces of paper - one with twenty-five numbered circles, randomly arranged, and the other containing thirteen numbers and twelve alphabets, also randomly arranged. Subjects must first draw lines between the circles in an ascending order, and then switch to drawing lines connecting the numbers and alphabets in sequential order (Ying Hui Tee, et al. 4).
   b. The time taken to complete both tasks is recorded, and the task-switching score is calculated by subtracting the time taken to complete the first part from the second part. In this case, a lower score denotes better performance.

III. Methodology

This is a multiple linear regression problem. In general, multiple linear regression assumes a linear relationship between several input variables x, and a single output variable y. Various approaches were taken in trying to build a successful predictive model, including preliminary inferential statistics, followed by basic and regularized linear regression.

A. Data Wrangling & Cleaning

First, any duplicate data were dropped. Twenty-seven values from four separate columns were missing. As such, the missing values from all columns were dropped and 486 observations – out of 513 - were retained for further exploration. No outliers were apparent.

The columns in the dataset are as follows:

Column Name	Column Description
Actual_age_C	Age
Sex	Sex
Weight	Weight (kg)
Height_m	Height (m)
BMI	BMI Score
BMI_for_age	Calculated BMI-for-Age Score
Physical_fitness_score	Physical Fitness Score
PFS_CAT	Physical Fitness Score Category 1. Poor 2. Low Average 3. High Average 4. Good 5. Excellent
Breakfast	Number of Breakfasts Consumed in a One-Week Period
Lunch	Number of Lunches Consumed in a One-Week Period
Dinner	Number of Dinners Consumed in a One-Week Period
PA_total_score	Physical Activity Score
PA_CAT	Physical Activity Category 1. Low PA Level

	2. Moderate PA Level 3. High PA Level
Sleep_weekdays	Average Hours of Sleep on Weekdays
Sleep_weekend	Average Hours of Sleep on Weekends
Sleep_percent	Sleep Percent
Global_sleep_CAT	Global Sleep Category 1. Poor sleep quality 2. Good sleep quality
Household_size	Household Size
Income_CAT	Income Category 1. RM < 2300 2. RM 2300 - RM 5599 3. RM > 5600
Edu_father	Father’s Level of Education 1. No formal education 2. Primary education 3. Secondary education 4. Tertiary education
Edu_mother	Mother’s Level of Education 5. No formal education 6. Primary education 7. Secondary education 8. Tertiary education
Interference_score_ALL	Interference Score
WM_total	Working Memory Score
CF_total	Cognitive Flexibility Score

Table 1. Column Names and Descriptions

B. Graphical Exploratory Data Analysis

An initial exploration of the dataset indicated that only seventeen students were 14 years old (3.5%) and thirty-three were 16 years old (6.8%). The spread of 12, 13, and 15-year-olds was relatively even. There were also 102 more responses from female students than male students.

Visual EDA followed and box plots were created for the categorical variables in the dataset which included sex, physical fitness category, physical activity category, global sleep category, income category, and household size. The results were as follows:

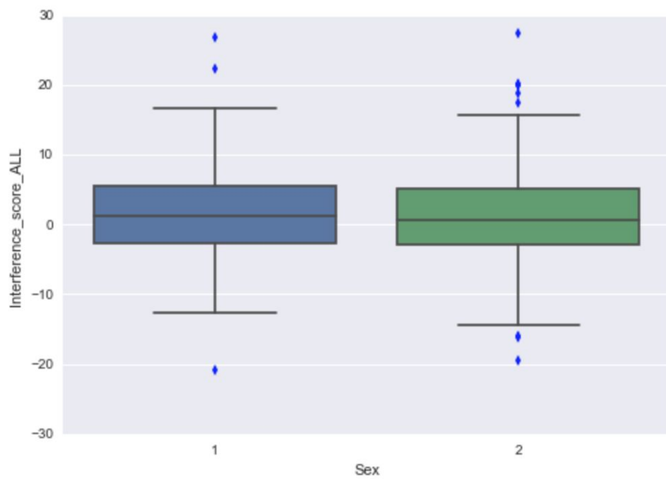


Figure 2-1. Sex vs. Interference. 1 = male; 2 = female.

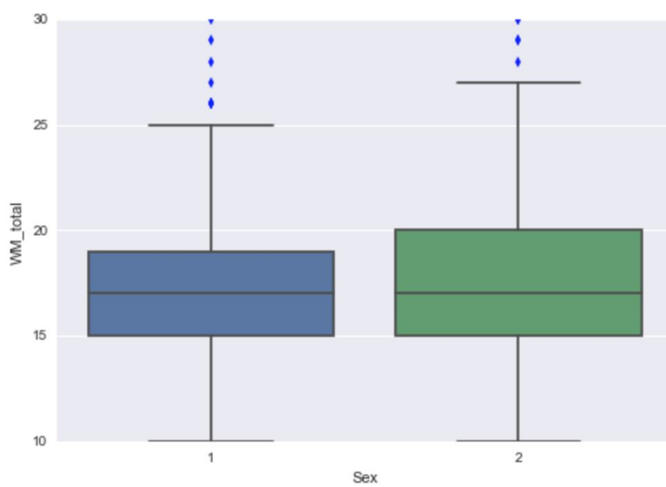


Figure 2-2. Sex vs. Working Memory. 1 = male; 2 = female.

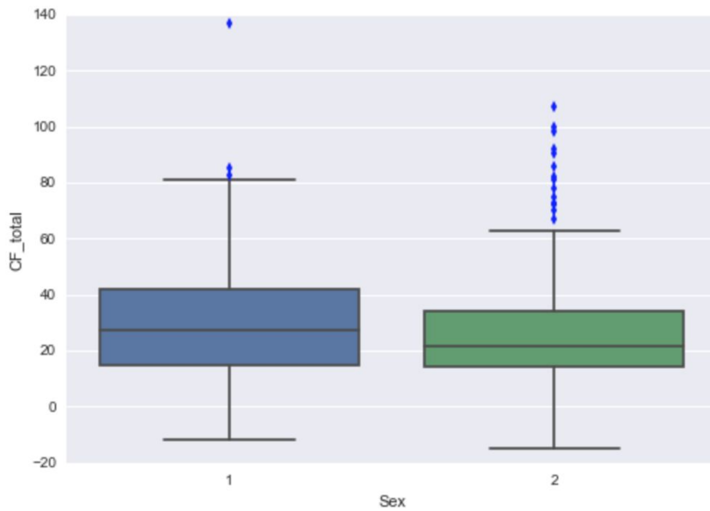


Figure 2-3. Sex vs. Cognitive Flexibility. 1 = male; 2 = female.

The interference distribution was very similar for males and females (Fig. 2-1). There was slightly more variation in the distribution for working memory, while it became clear that there is a significant difference between males and females when it comes to cognitive flexibility (Fig. 2-2, 2-3).

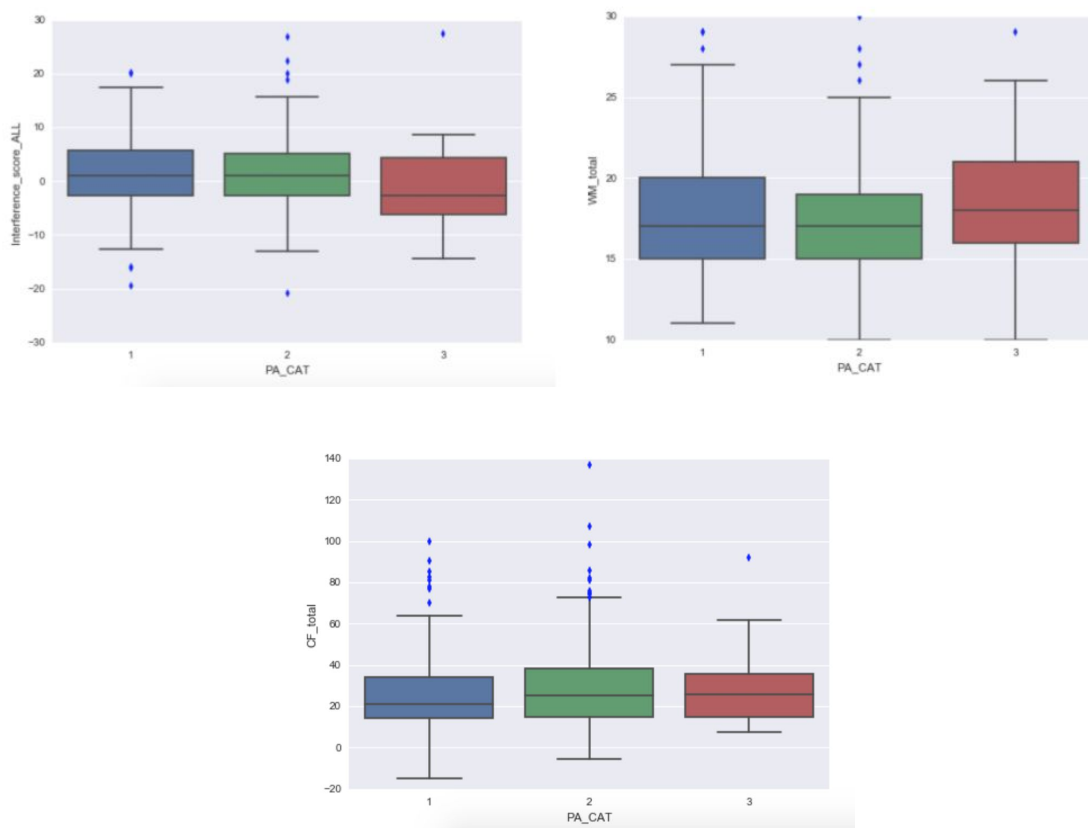


Figure 3. Physical Activity Category vs. Executive Function. 1 = low; 2 = moderate; 3 = high.

Similarly, the executive function distributions were similar for all physical fitness score categories, though the differences in physical activity distributions suggested there are differences between groups (Fig. 3).

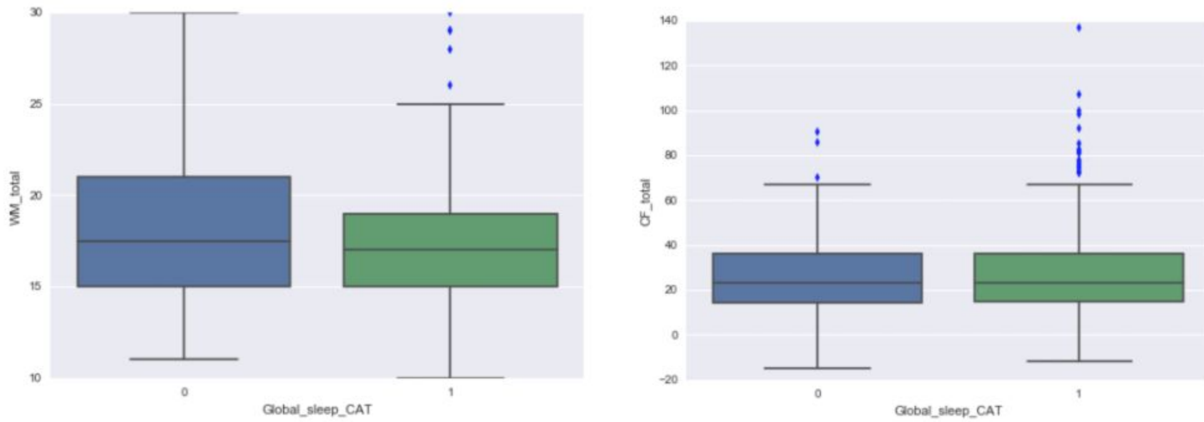


Figure 4. Global Sleep Category vs. Executive Functions (Working Memory and Cognitive Flexibility). 0 = poor sleep quality; 1 = good sleep quality.

Global sleep category is divided into two groups in the dataset: poor sleep quality and good sleep quality. The working memory median was the same for both sleep category groups, but there was a large difference in their distributions (Fig. 4). There was no variation for cognitive flexibility.

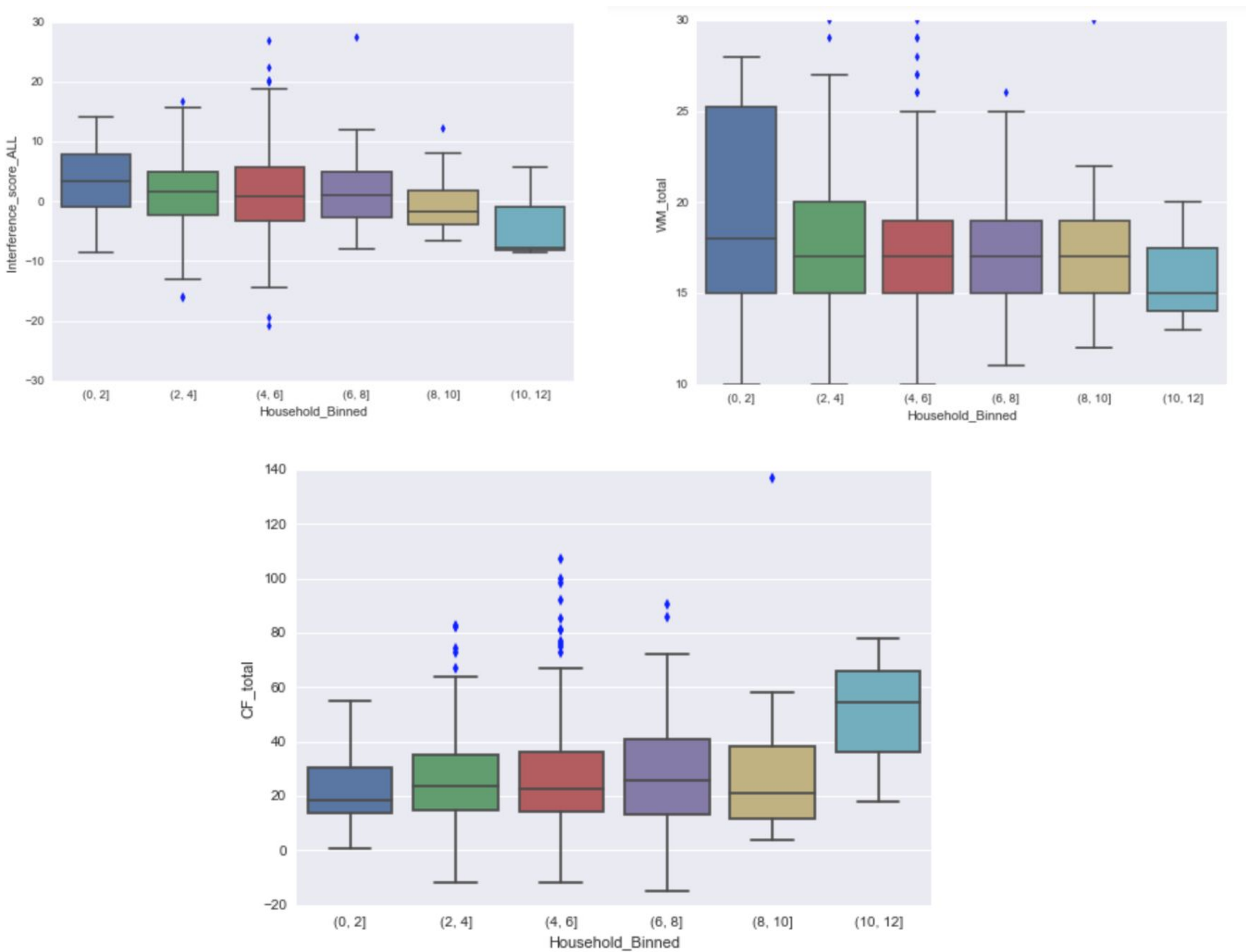


Figure 5. Household Size vs. Executive Functions.

Household size showed large discrepancies for all executive functions (Fig. 5).

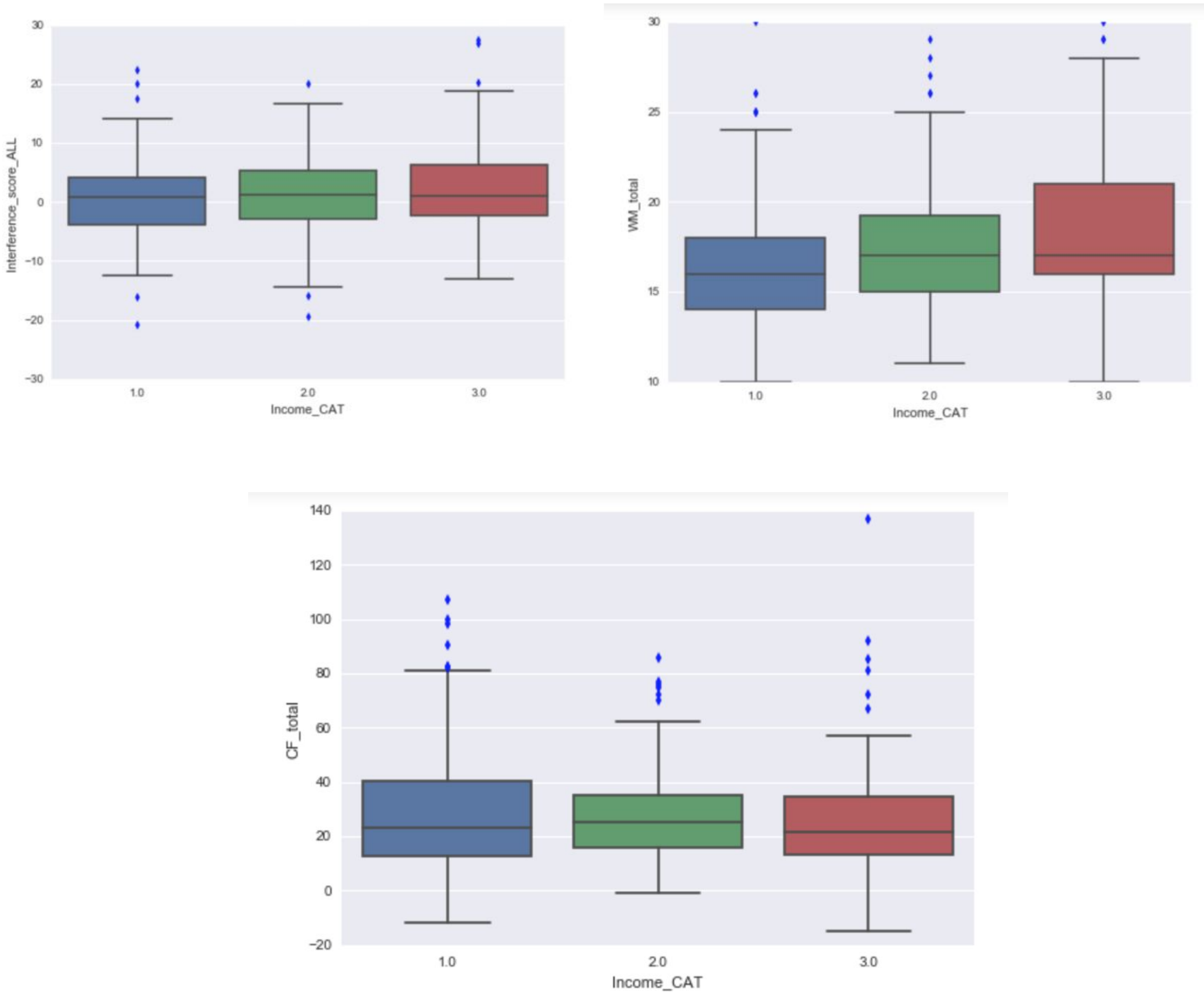


Figure 6. Monthly Household Income Category vs. Executive Function.
1 (RM < 2299); 2 (RM = 2300 - 5599); 3 (RM > 5600)

Monthly household income category was split into three groups: RM < 2299, RM = 2300 - 5599, and RM > 5600, where RM stands for the Malaysian Ringgit, the country's official currency.

For reference, the annual salary for each group (in dollars) is as follows:

- Group 1 (RM < 2299) = USD < $6,760.00
- Group 2 (RM = 2300 - 5599) = USD = $6,763.00 - $16,462.00
- Group 3 (RM > 5600) = USD > $16,465.00

Income category showed a stable distribution for interference and cognitive flexibility, but displayed a strong discrepancy when it came to working memory (Fig. 6).

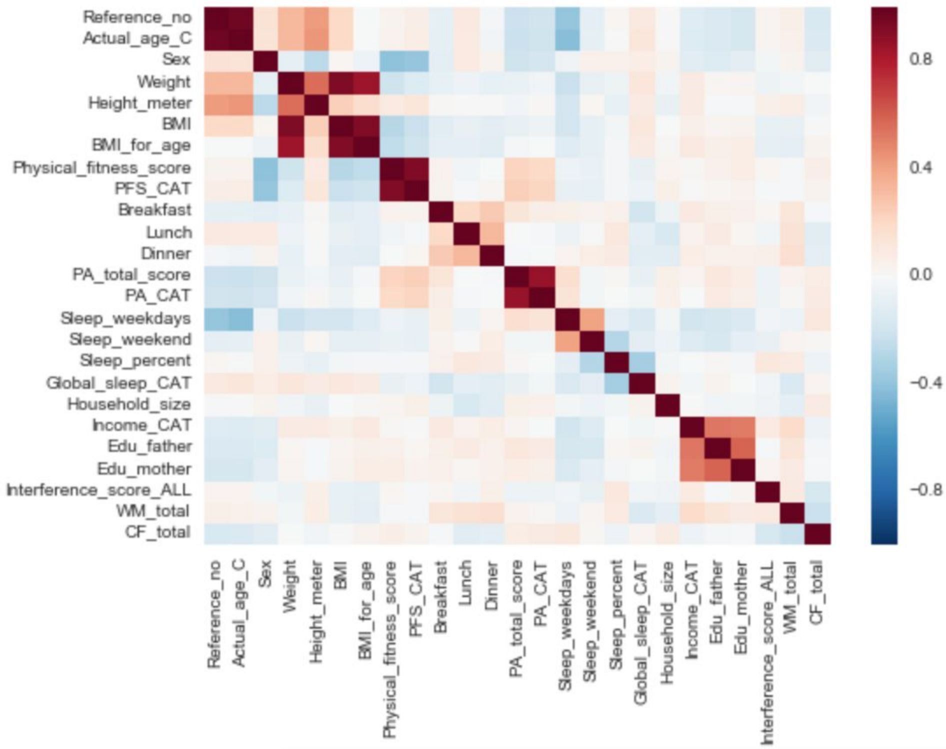


Figure 7. Standard correlation matrix for all variables.

Finally, a standard correlation matrix was graphed in an effort to find noticeable correlations and check for collinearity (Fig. 7). This was followed by inferential statistical analysis to determine whether any perceived correlations were statistically significant.

C. Inferential Statistics

Two-sample t-tests were conducted for two categorical variables - sex and global sleep category. The p-value of 0.010 was calculated for sex with regard to cognitive flexibility (CF), leading to the rejection of the null hypothesis at the 0.05 significance level, signifying that mean CF scores are statistically different for males and females. A Bonferroni correction was performed on the sample, which gave a corrected significance value of 0.03, and reiterated the need to reject the null hypothesis in favor of the alternate.

The null hypothesis was also rejected for global sleep category with working memory (p-value = 0.0007; Bonferroni α = 0.0021), 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 features, all of which had three or more categories within their respective variables. The One-Factor ANOVA was the most appropriate parametric test for determining statistical significance – however, the physical fitness and physical activity categories failed to meet the proper ANOVA conditions. 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.

```
In [104]: # Bonferroni Correction

          p_income = [0.388, 0.047, 0.252]

          p_adjusted_income = multipletests(p_income, alpha=0.05, method='bonferroni')
          p_adjusted_income

Out[104]: (array([False, False, False], dtype=bool),
           array([ 1.    ,  0.141,  0.756]),
           0.016952427508441503,
           0.016666666666666666)
```

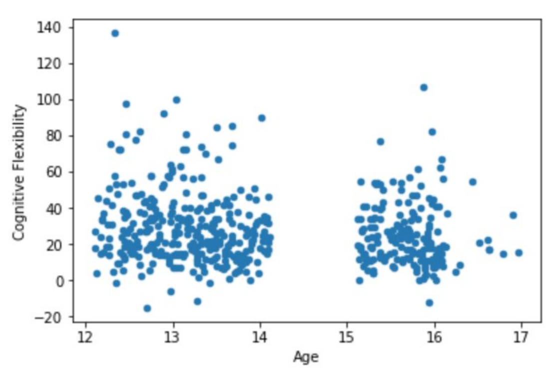
Figure 8. Bonferroni Correction for Income Category with Interference

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 (p-value = 0.047 at ɑ = 0.05). The Bonferroni correction produced a corrected ɑ = 0.141 but the output stated that the null hypothesis could not be rejected (Fig. 8).

The income category/working memory and income category/cognitive flexibility parameters met the conditions for ANOVA and the following results were obtained:

1) There is a significant difference in mean working memory scores between students from low-income and high-income backgrounds
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 (Fig. 9 below):



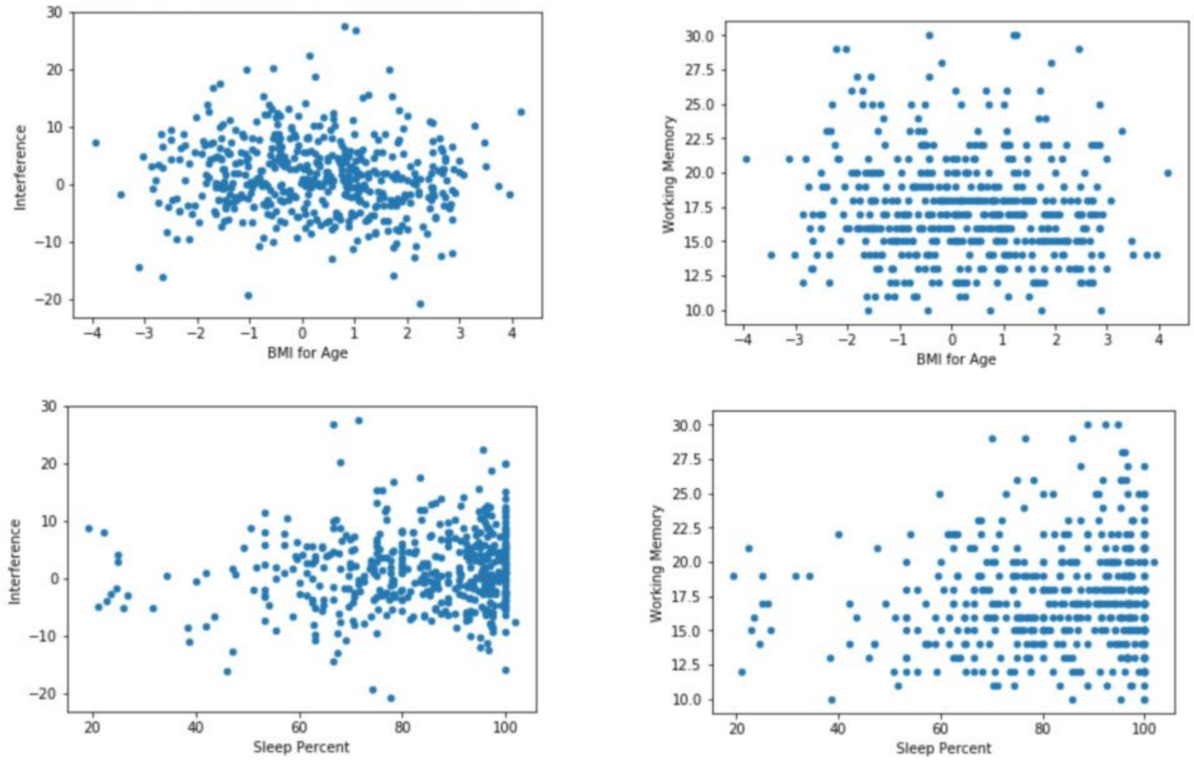


Figure 9.Graphical Results of Pearson Correlation

- **Age:** There is a significant negative correlation between age and cognitive flexibility ($r = -0.146$; $p = 0.012$)

BMI-for-Age: There is a significant negative correlation between BMI-for-age and interference ($r = -0.089$; $p = 0.048$), as well as BMI-for-age and working memory ($r = -0.094$; $p = 0.038$). The correlation between BMI-for-age and cognitive flexibility is insignificant ($r = 0.038$; $p = 0.396$).

- **Sleep_percent:** There is a significant positive correlation between sleep percent and interference ($r = 0.110$; $p = 0.015$), as well as sleep percent and working memory ($r = 0.092$; $p = 0.042$).

- **Household_size:** There is a significant negative correlation between household size and working memory ($r = -0.098$; $p = 0.0306$).

D. Model Fitting & Feature Selection

Linear Regression (Sci-Kit Learn)

Sci-Kit Learn provides an algorithm to run simple linear regression models using train-test splits. Its output is R^2, which denotes the “goodness of fit” of the model based on its sums-of-squares. R^2 describes the proportion of variance of the dependent variable explained by the regression model.

Basic linear regression was conducted three separate times – once for each of the three target variables (interference, working memory, and cognitive flexibility). Each train-test split included all predictor variables, trained against one executive function score. Predictions were made on the test data and the outcomes were as follows:

- Interference R^2 = -0.029
  - 5-Fold CV Score = -0.04

- Working Memory R^2 = 0.11
  - 5-Fold CV Score = 0.011

- Cognitive Flexibility R^2 = -0.023
  - 5-Fold CV Score = -0.13

Based on the negative R^2 and cross-validation scores, it became clear that feature selection was necessary.

GridSearching

GridSearching, a form of hyperparameter optimization, was performed in an effort to determine the ideal parameters for three types of regularized regression – LASSO, Ridge, and Random Forest Regression. The following parameters were found:

LASSO (ɑ) = 0.1
Ridge (ɑ) = 1000
Random Forest (max_depth) = 2

LASSO Regression

LASSO regression for each target variable followed next, with ɑ = 0.1 and the same train-test splits used for basic linear regression.

LASSO is a very useful linear regression model which focuses on shrinkage estimation in an effort to minimize collinearity. It performs L1 regularization, adding penalties to the absolute value of the magnitude of coefficients. As such, insignificant coefficients are eliminated from the model entirely.

The following LASSO scores were generated:

- Interference = -0.0102
- Working Memory = 0.111
- Cognitive Flexibility = -0.425

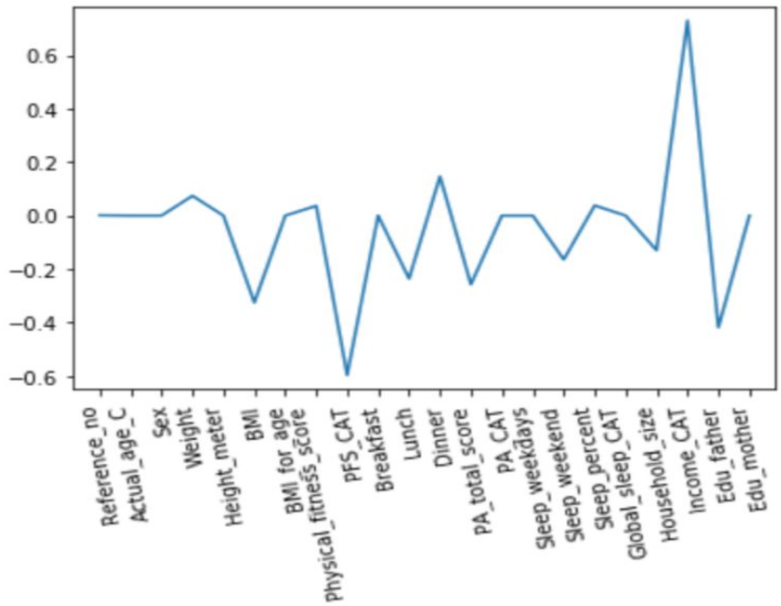


Figure 9. LASSO Regression (Interference)

The LASSO regression result for Interference was graphed, and it became clear that 'PFS_CAT' and 'Income_CAT' are the two most important features when predicting Interference scores (Fig. 9).

Another basic linear regression analysis for Interference followed, this time with only the two variables included as features. The model showed a marked improvement, with R^2 = 0.018 and 5-Fold CV Score = 0.003.

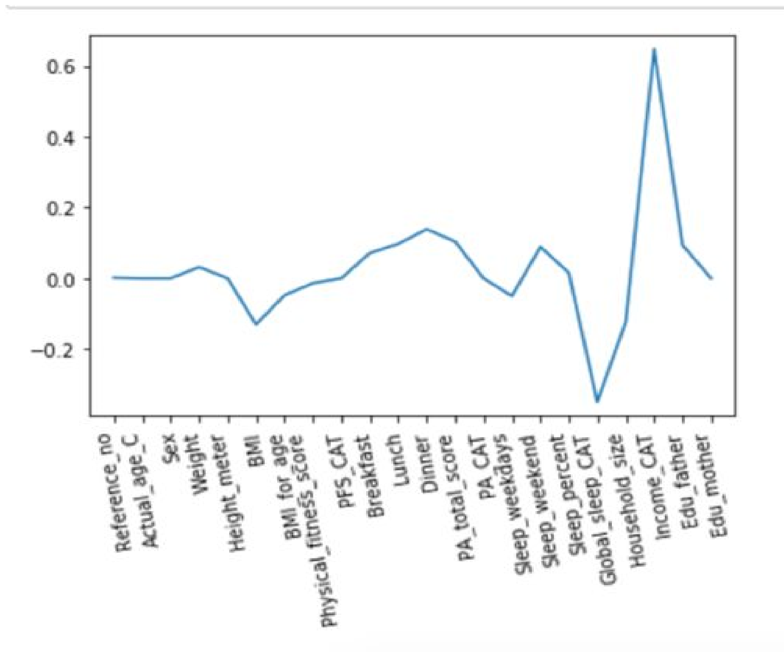


Figure 10. LASSO Regression (Working Memory)

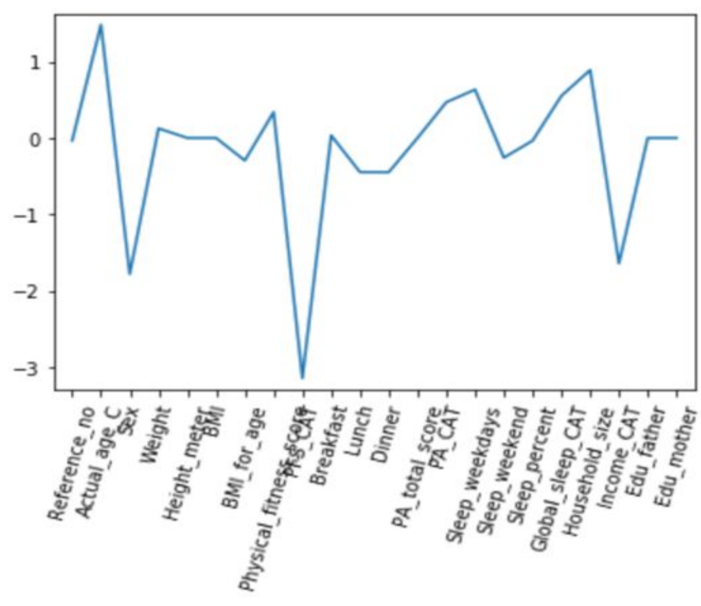


Figure 11. LASSO Regression (Cognitive Flexibility)

The LASSO regression results for Working Memory and Cognitive Flexibility were also graphed (Fig. 10-11). 'Global_sleep_CAT' and 'Income_CAT' were the two most salient features for predicting Working Memory scores, whereas 'Actual_age_C', 'Sex', 'PFS_CAT', 'Household_size', and 'Income_CAT' were the features that best predicted Cognitive Flexibility scores. The basic linear regression analyses for both targets improved, with WM R^2= 0.094 and 5-Fold CV Score = 0.043, and CF R^2 = 0.053. The 5-Fold CV for CF performed worse however, with CV = -0.039.

Each LASSO regression was followed by linear regression using Ordinary Least Squares (OLS) in StatsModels, with LASSO-selected features for updated train-test splits.

Linear Regression (StatsModels)

Linear regression via StatsModels allows for model fitting using OLS. OLS is a linear least squares method for estimating the coefficients of a linear model by minimizing the squared residuals. Statsmodels provides detailed statistical information not available through Sci-Kit Learn.

The outputs of this algorithm include R^2 and an adjusted R^2 , the latter of which was used in this analysis. The adjusted R^2 takes the number of variables in the model into account in an effort to minimize any artificial inflation of the R^2 value due to increased dimensionality.

- Interference $R^2 = 0.025$
    - Adjusted $R^2 = 0.014$

- Working Memory $R^2 = 0.087$
    - Adjusted $R^2 = 0.074$

- Cognitive Flexibility $R^2 = 0.047$
    - Adjusted $R^2 =0.036$

Hierarchical Regression

Finally, hierarchical linear regression was performed, three separate times. Hierarchical regression is a type of OLS in which predictor variables are added or removed from a model in phases, in an effort to increase the viability of the model.

The protocol followed the one detailed in the journal article so that the R^2 values could be compared fairly. The model was split into three distinct phases:

- Phase I: Age, sex, monthly household income, parents' years of education, and household size
- Phase II: BMI-for-age
- Phase III: Meal intakes, sleep quality, physical activity and aerobic fitness
- Level of Significance: $p < 0.05$

Then, linear regression models were run using both Sci-Kit Learn and StatsModels. The following comparisons were made:

Overall Interference R^2 = 0.033 (journal); 0.082 (sklearn); 0.023 (StatsModels)
Overall WM R^2= 0.128 (journal); 0.040 (sklearn); 0.081 (StatsModels)
Overall CF R^2 = 0.061 (journal); 0.088 (sklearn); 0.035 (StatsModels)

Ridge Regression

Ridge is another form of regularized regression in which the issue of collinearity is aided by adding a degree of bias to regression estimates.

The following Ridge scores were generated with the features included in Phase III of hierarchical regression:

- Interference = 0.013
- Working Memory = -0.052
- Cognitive Flexibility = 0.021

Random Forest Regression

Random Forest performs regression by implementing multiple decision trees and bootstrap aggregation.

The following Random Forest scores were generated with the features included in Phase III of hierarchical regression:

- Interference = -0.032
- Working Memory = -0.037
- Cognitive Flexibility = -0.086

IV. Conclusion

In conclusion, I do not believe that this dataset is suitable for building functional or accurate predictive models for multiple reasons.

Visual EDA and statistical inference dictated that the following features were most influential when predicting each EF score:

- **Interference:** Household size, income category, BMI-for-age, and sleep percent
- **Working Memory:** Global sleep category, household size, income category, BMI-for-age, and sleep percent
- **Cognitive Flexibility:** Sex, household size, and age

Linear regression models were built and cross-validated, and corroborated these results. The following features showed significance:

- **Interference:** Income category, BMI-for-age, and sleep percent
- **Working Memory:** Age, income category, BMI-for-age, and global sleep category
- **Cognitive Flexibility:** Age

Statistically, BMI proved to factor strongly in predicting executive function abilities. Regardless of statistical significance however, the R^2 values obtained through each run of linear regression were very low.

A basic linear regression in Sci-Kit Learn was followed by LASSO regression, and the viability of each model improved significantly as the most important features for each of the three executive EF scores were selected when making train-test splits. The subsequent StatsModels regression runs provided similar or slightly improved R^2 values.

Finally, hierarchical regression for each factor of executive function was conducted. The R^2 found using both Sci-Kit Learn and Statsmodels were similar or slightly higher than the ones published in the journal article. Nevertheless, the R^2 values themselves were very low.

Cross-validation was also unsuccessful as many of the 5-fold CV scores were actually negative.

Thus, even though the results could be deemed statistically significant, the dataset could not help draw results of practical significance.

V. Client Recommendations

Though the viability of all linear regression models were deemed insignificant, some client recommendations can still be made. For one, general practitioners can use a patient's BMI and the average of their EF scores to determine whether a psychiatric referral is needed for the diagnosis and treatment of ADHD, or any comorbid mental disorders, i.e. anxiety or depression.

On the other hand, if a patient only shows certain risk factors, doctors can recommend Cognitive Behavioral Therapy (CBT) to help patients develop better self-regulation skills. At home, parents and the students themselves can use the data to implement better daily routines.

VI. References

"Stroop Color Word Test." *BBC - The One Show*, 17 Feb. 2009, www.bbc.co.uk/blogs/theoneshow/onepassions/2009/02/body-tricks-stroop-test.html.

Ying Hui Tee, Joyce, Wang Ying Gan, Kit-Aun Tan, and Yit Siew Chin. "Obesity and unhealthy lifestyle associated with poor executive function among Malaysian adolescents." *Public Library of Science,* vol. 13, no. 4, 2018, https://doi.org/10.1371/journal.pone.0195934