Progress Updates -Team 5

Project Recap

Based on our literature review and data analysis of the UK Biobank data, we found there was negative association between screen time and diet quality, and a positive relationship between diet quality and health outcomes. Also, as patients get older, they tend to spent more time in front of television or computer. Moreover, limited study concerns about the screen time impact on diet quality among older people. The goal of our project is to explore the relationship between screen time and diet quality, and how diet quality further affects overall health among older adults aged 65 and older. To explore the idea, we propose the following hypothesis and specific aims:

> **Hypothesis.** Among older adults aged 65 and older (65+), increased daily screen time will impact their diet behaviors and impair their diet qualities, which further causes adverse health outcomes.
>
> **Specific Aim 1.** To investigate the association between time spent on television or computer among older adults and their dietary quality.
>
> **Specific Aim 2.** To investigate the association between dietary quality and overall health among older adults aged 65+.

Updates of Diet Quality Indexing

We proposed to use existing diet quality index to calculate a diet quality score, which required us to calculate food groups variety, protein sources variety, Adequacy, Moderation, and Balance respectively. Unfortunately, the data we have only provides the frequency of food intake, where we are not able to get the exact quantity of food intake or find out the exact nutrient value for some foods like cooked vegetables. Therefore, we decided to employ a new method based on

Mediterranean Diet Score, which stratifying by sex, assigns 1 to intake above median score and zero otherwise. After that, we combined Mediterranean Diet Score, Food groups variety, and Protein sources variety with a reasonable set of weights to get the final dietary score for each observation.

Updates of Modeling

To investigate the relationship between diet score and overall health rating in specific aim 2, we planned to construct a random forest model together with some baseline characteristics. After three rounds of sieve, we selected the following features from 324 covariates: Standing height, Seated height, Number of days a week walked 10 minutes, Number of days a week of vigorous physical activity 10 minutes, Morning evening person chronotype, Sleeplessness, Frequency of tiredness, Family friend visit, and Trunk fat percentage. We first removed the variables with too many missing values, or not related to our study. In second round, we filtered out the features with extremely high correlation of the overall health ratings or with a variance smaller than 0.05. Finally, we used multinomial logistic regression to extract the essential covariates.

We also finished writing codes on the simulated data. We built random forest classifiers for four different levels: poor, fair, good, and excellent. Then the hyper-parameters were tuned based on the F-1 score per class in a one-vs-rest manner. Eventually, we would expect an acceptable F-1 score ($> 60\%$) for the calculations of counterfactuals to be eligible.