ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #36201

Investigation of Dominant Daily Uptake Factors on Gut Health from Samples in the Database of National Health and Nutrition Examination Survey

Margaret Dugoni, Villanova University
Ms. Nicola Iris Kaye, Villanova University

3rd year Chemical Engineering major with a minor in Biochemical Engineering.

Dr. Zuyi (Jacky) Huang, Villanova University

Zuyi (Jacky) Huang is an Associate Professor in the Department of Chemical Engineering at Villanova University. He teaches Chemical Process Control (for senior students) and Systems Biology (for graduate students) at Villanova. He is enthusiastic in applying innovative teaching methods in class to educate students with modeling and control skills. His research is focused on developing advanced modeling and systems analysis techniques to manipulate microbial biological systems for generating biofuels from wastewater and for combating biofilm-associated pathogens. His BESEL group developed the first model for microbial desalination cells and the first metabolic modeling approach for quantifying the biofilm formation of pathogens.

Investigation of Dominant Daily Uptake Factors on Gut Health from Samples in the Database of National Health and Nutrition Examination Survey

Margaret Dugoni*, Nicola Kaye*, Zuyi (Jacky) Huang
Department of Chemical and Biological Engineering, Villanova, PA, 19085
*- equal contribution

Abstract

The diversity of healthy gut bacteria in the human digestive system is linked to positive health effects. The intestinal microbiome, including but not limited to, *Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, Verrucomicrobia,* and *Fusobacteria*, can aid immune system, improve digestion, and support mental health. Since caffeine, sugars, and alcohol are frequently consumed and may upset the environment of the gut microbiota, it is important to investigate how these factors influence gut health. An unhealthy gut can lead to several negative effects on health, such as indigestion issues, low sleep quality, and poor mental health. The hypothesis that drives this research is that individuals who consume higher than the recommended volume of sugar, alcohol, or caffeine will experience negative effects on their gut health, which can then impact their digestion, mental health, and quality of sleep. To investigate this hypothesis, data from the database of National Health and Nutrition Examination Survey (NHANES) was thoroughly screened for samples containing all factors studied in this work, including the daily consumption of sugar, alcohol, and caffeine. Levels of C-Reactive Protein (CRP) and Tissue Transglutaminase IgA were two variables also extracted from NHANES that indicate gut health as they are involved in gut inflammation. The NHANES data were imported into the R programming platform, which was used to run a dominance analysis on all factors to examine the largest influence on gut inflammation. Principal component analysis and hierarchical clustering were then performed to examine the variance of the variables as well as the grouping. Finally, a t-test was conducted to examine if consuming higher than the recommended value of caffeine, sugars, or alcohol has a significantly different effect on gut health than consuming the recommended value. The t-test was able to further validate the significance of the identified dominant factors. The dominance analysis demonstrated that the total amount of consumed sugar has the greatest impact on C-Reactive Protein while consuming more caffeine had the largest impact on Tissue Transglutaminase IgA. The average amount of alcohol also appeared to impact C-Reactive Protein. This work initiates the effort to examine overall gut health, which requires analysis on more variables including nutrients, diet, and human bowel health.

Key words: Dominance analysis, t-test, Gut health, NHANES, C-Reactive Protein, Tissue Transglutaminase IgA

Introduction

The diversity and volume of healthy gut bacteria in the human digestive system are linked to positive health effects. The intestinal microbiome—*Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, Verrucomicrobia, and FusobacteraiI*— can aid in the immune system, improve digestion, and support mental health [1][2]. There are several conditions in which the bacteria can live; however, this is an ideal pH and concentration of species [3][4]. The consumption of caffeine, sugars, and alcohol can upset the environment of the gut microbiota; an unhealthy gut can lead to several negative effects on health such as indigestion, poor sleep quality, and poor mental health

[2], [4-7]. The research group hypothesized that individuals who consume higher than the recommended volume of sugars, alcohol, and caffeine will experience negative effects on their health.

A study conducted by Lee analyzed the impact of drinking alcohol on gut microbiota [6]. The study collected data from many other articles that researched how alcohol can affect bacterial growth in the human intestine. One positive result they concluded was that alcohol can alter the gut in a way that prevents aging. They also concluded that nearly all alcohol-related liver disease was correlated with a lesser diversity in gut flora. The lesser gut flora diversity there is, the less supported the immune system is. One purpose of this project is to analyze how alcohol consumption can negatively affect bacterial growth, and in turn, how that affects human health.

Gonzalez and their team studied how coffee or caffeine consumption can affect gut flora growth [7]. The researchers grouped their participants into three groups: non, moderate, or high consumers of coffee. Each participant's sample was measured for microbiome health using gas chromatography. Though the research group did not find any significant results, they observed that the ratio of bacteria species was different amongst the type of coffee consumer. One purpose of the experiment outlined in this report is to analyze how caffeine consumption affects human health.

Data Analysis and Database

R is a free program commonly used in biochemical data analysis. This program is user-friendly and pairs with MATLAB in the academic setting [8]. In this project, data was imported from the NHANES database. Several statistical tests were coded and performed using R. The tests used in this experiment are dominance analysis, hierarchical clustering, principal component analysis, linear regression, and t-test.

A dominance analysis is used to determine the importance each considered variable will play in a regression. After determining the most important variable on the system, additional factors can be included to evaluate how the model changes. Dominance analysis is based on linear regression. A linear regression relates actual data (e.g., the daily consumption of sugar, alcohol, and caffeine) to predicted results (e.g., C-Reactive Protein concentration). The variables and their errors can be fit with a trendline. How well the trendline fits the data can be determined from the R^2 value. The value of R^2 indicates what percentage of the information in the model is accounted for by the trendline and its equation; an R^2 value close to one indicates a good fit [9]. As for dominance analysis, the enhancement of R^2 value is used as the indicator of the significance of the added variable into the linear regression model. The variables are then ranked on the basis their influence of the R^2 value. A few forms of dominance can occur; complete dominance is when one variable completely influences the subject, regardless of the number of additional variables. Conditional dominance occurs when a certain variable is important in a different part of the model. Finally, general dominance is when the average contribution of a variable is generally larger than the rest of the variables [9].

While dominance analysis is able to rank the variables, it does not illustrate relationship between those variables. To address this, Principal Component Analysis (PCA) is used to reduce the dimensions of a large set of data so that hierarchical clustering can be used to illustrate the similarity of those variables in their influence on the selected output. In particular, PCA allows the data to be grouped in terms of the variance it contributes to the subject [10]. Each point on the PCA has a projection or score, which comes from the original data set. Typically, only graphed is principal component 1 (PC1) and principal component 2 (PC2). The grouping can be complemented by the hierarchical clustering， which is a diagram that depicts related variables in a tree structure [10]. It is used when the researcher does not know how many groups of data they want. The vertical distance between the variables indicates how similar they are; a large distance indicates little similarity.

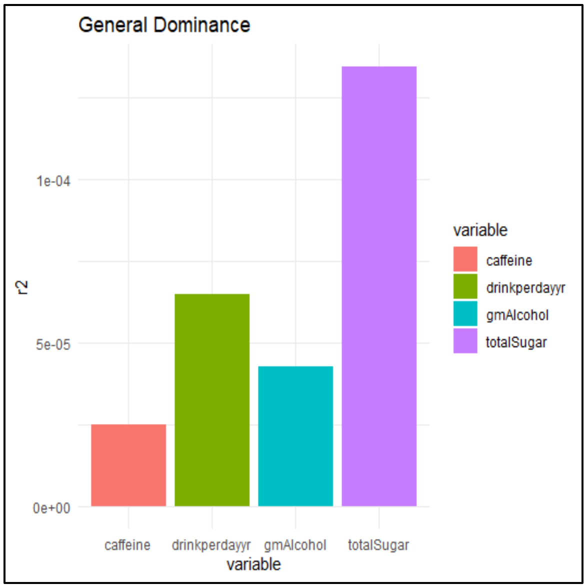
The statistical t-test can be either a single or double parameter test; this experiment used two groups. The t-test compares the means of two independent groups. This test outputs the p-value, which can be used to understand statistical significance. If the test produces a p-value of greater than 0.05, then there is no significant difference in the mean values of the groups compared. If the test produces a p-value of less than 0.05, then there is a significant difference in the means of the groups, signifying meaningful results [10]. In this work, t-test was used to compare the groups selected based on a specific variable to investigate the significance of that variable on influencing the output. The results from t-test were further used to validate the significance of the variables identified from dominance analysis.

This work utilized the Center for Disease Control's database, NHANES, to collect various data on gut bacteria and physical health. Data was collected data for variables including alcohol use, caffeine consumption, sugar consumption, and gut inflammation. Using this data, the research group performed thorough statistical analysis. First, a dominance analysis was performed to determine the variables the most highly influenced CRP and Tissue Transglutaminase IgA. Principal component analysis, hierarchical clustering, and linear regression were then completed to better visualize the variables. Finally, a t-test was completed to see if the consumption of a higher than the recommended value of caffeine, sugars, or alcohol has a significantly different effect on each of the selected outputs than a recommended value. It was hypothesized that those who consume a higher than the recommended value of alcohol, caffeine, or sugar will exhibit negative effects on their health.

Results and Discussion

The first parameter studied was the level of CRP, which is more present when the gut is inflamed. The dominance analysis showed that the factor that most influenced the presence of CRP was sugar consumption, shown in Figure 1. The hierarchical clustering, Figure 2, showed the caffeine, sugar, and alcohol consumption related to CRP. Grams of alcohol consumed related most closely to CRP. PCA, Figure 3, was used to show how similarly each of the studied variables were linked. Alcohol consumption and CRP were related most similarly in the space characterized by PC1 and PC2. A linear regression was generated to view the real vs predicted value of sugar, the variable that most influenced the presence of CRP, Figure 4. The value of R^2 was 0.49. Finally, a two-sample t-test was performed to see if the mean presence of CRP was significantly different between groups that consumed either above or below the recommended 24 grams of sugar per day [11]. The test produced a p-value of 0.0041, which supports the alternate hypothesis. The mean value of CRP present is significantly greater in those who consume greater than recommended values of sugar than those who consume at or below the recommended dosage.

A similar statistical analysis was conducted for a blood test variable that correlates with gastrointestinal tract disorders: Tissue Transglutaminase IgA. The presence of increasing tissue transglutaminase antibodies in the body indicates a gut disorder or diseases such as chronic inflammation or Celiac disease. The dominance analysis, Figure 5, showed that the factor that most



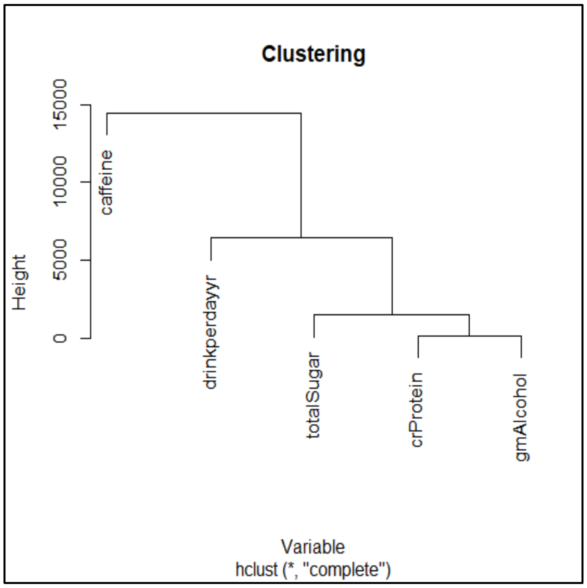
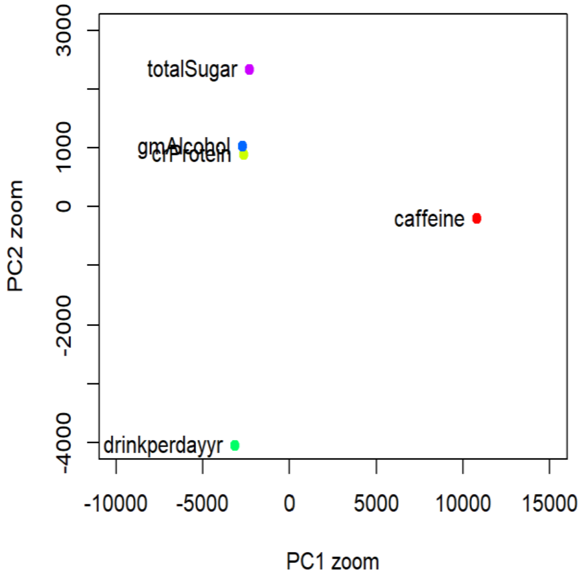


Figure 1. (Left) General dominance for C-reactive protein. The dominant input is total sugar.
Figure 2. **(Right)** Hierarchical clustering that demonstrates similarities between CRP and grams of alcohol consumed. Also related is total sugar, drinks per year, and caffeine consumption.



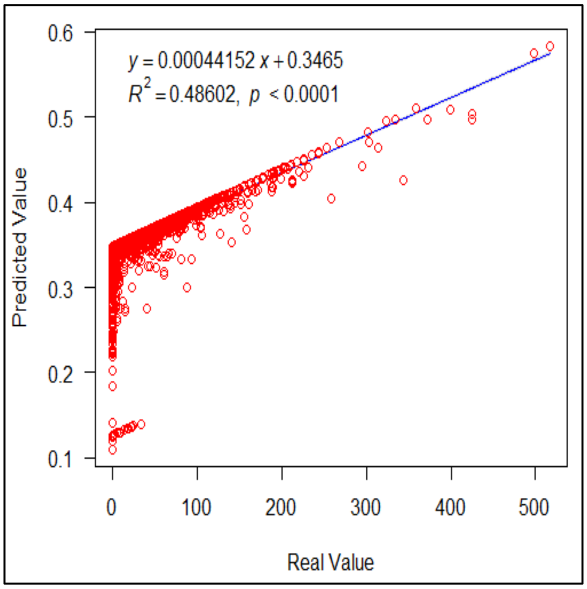
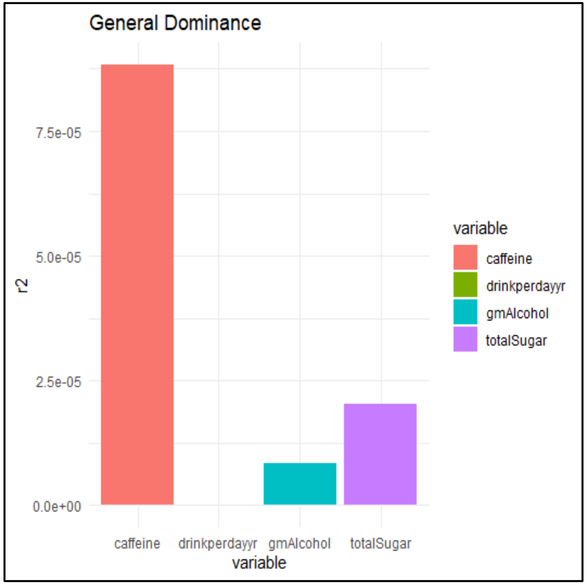


Figure 3. (Left) Principal component analysis of inputs on C-reactive protein. CRP and alcohol consumption are related most similarly in PC1 and PC2.
Figure 4. **(Right)** Linear regression of the dominant variable, total sugar, on CRP.

impacted Tissue Transglutaminase IgA was caffeine consumption. The hierarchical clustering, Figure 6, showed the caffeine, sugar, and alcohol consumption related to Tissue Transglutaminase IgA. Grams of alcohol consumed per day related most closely to Tissue Transglutaminase IgA. PCA, Figure 7, was used to show how similarly each of the studied variables were linked. Alcohol consumption and Tissue Transglutaminase IgA were related most similarly in PC1 and PC2. A linear regression was generated to view the real vs predicted value of caffeine, the variable that most influenced the presence of Tissue Transglutaminase IgA, Figure 8. The value of R^2 was 0.76.

Finally, a t-test was run to determine if there was a significant difference in the mean level of Transglutaminase IgA present in those who consume either above or below the recommended 135mg of caffeine per day [12]. The p-value was 0.16, indicating that the null hypothesis was supported. There is no significant difference in the mean volume of tissue antibodies in those who consume either above or below the recommended amount of caffeine. This may be due to the relatively low R^2 contribution of caffeine to the output (i.e., Transglutaminase IgA), as shown in Figure 5.



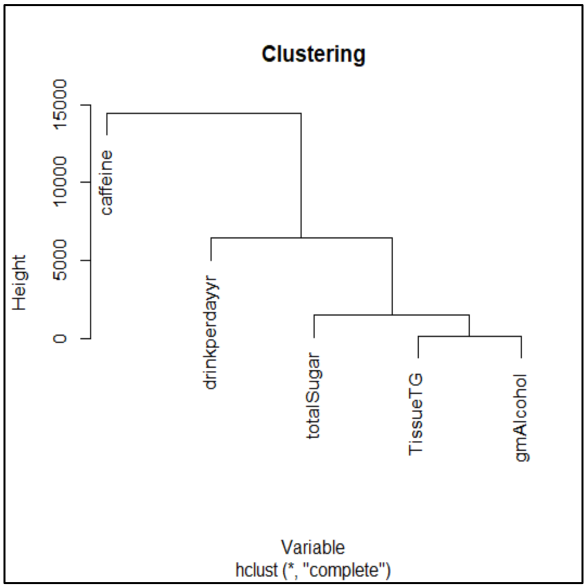
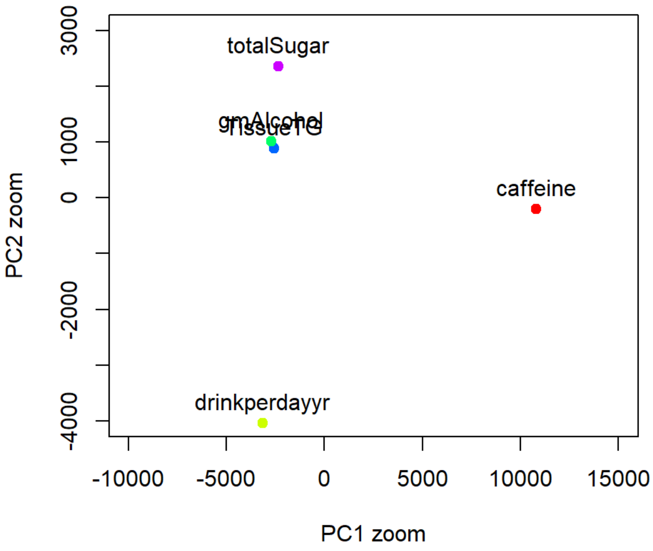


Figure 5. (Left) General dominance for Tissue Transglutaminase IgA. The dominant input is caffeine consumption.
Figure 6. **(Right)** Hierarchical clustering that demonstrates similarities between Tissue Transglutaminase IgA and grams of alcohol consumed. Also related are total sugar consumer, drinks per day per year, and caffeine consumed.



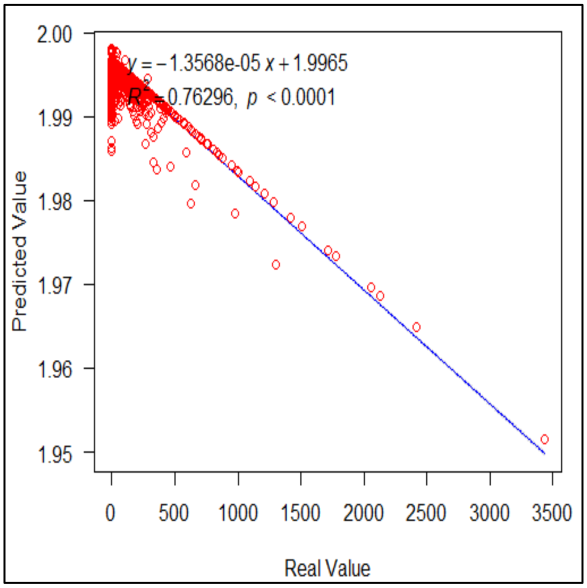


Figure 7. (Left) Principal component analysis of inputs on Tissue Transglutaminase IgA. Tissue Transglutaminase IgA and alcohol consumed per day are related most similarly in PC1 and PC2.
Figure 8. **(Right)** Linear regression of the dominant variable, caffeine consumption, on Tissue Transglutaminase IgA.

Discussion

It is interesting to note how both analyses showed that the most dominance variable was not the variable that was most similarly grouped. It is possible that the consumption of one item alone is not enough to induce gut inflammation. In other words, the dominance analysis depends on the R^2 contribution of variables in all possible combinations of linear regress models, while PCA and clustering mainly focus on the dynamic similarity shown in the data. The mixture of sugar, alcohol, and caffeine consumption could be studied to understand their impact more accurately on intestinal health. The effort should also be made to examine overall gut health more broadly, which could require analysis on more variables including nutrients, diet, and human bowel health. Finally, finding variables that more directly indicate gut health could help to hone the understanding of how the diet impacts other aspects of human health.

This project shows an example of how to integrate national databases into engineering education, especially via term projects. In addition to the database of National Health and Nutrition Examination Survey, there are various databases freely available on CDC/NIH website. These are valuable resources for engineering educators. Since these databases are free, students with computers can get involved in designed projects so that more inclusive and diverse workforce with engineering skills may be developed for the future in the US. In addition, this project integrates disciplines of biology, data analysis, engineering, and nutrients. It thus shows an example of multi-disciplinary engineering project. On the basis of this project, the instructor (Huang) is developing a STEM outreach program with a local school district. More quantitative feedback comments will be collected and provided by the instructor in the future.

Conclusion

The purpose of this project was to analyze if a higher than recommended consumption of sugars, alcohol, or caffeine would have a significant effect on human health as observed by C-Reactive Protein and Tissue Transglutaminase IgA. A dominance analysis showed that sugar consumption most greatly impacted the presence of CRP. The clustering and PCA showed that CRP and alcohol consumption were related most similarly. The linear regression relating CRP and sugar consumption had an R^2 of 0.49, and the t-test had a p-value of 0.004. The mean value of CRP present is significantly greater in those who consume greater than recommended values of sugar than those who consume at or below the recommended dosage. A dominance analysis showed that caffeine consumption most greatly impacted the presence of Tissue Transglutaminase IgA. The clustering and PCA showed that alcohol consumption and Tissue Transglutaminase IgA were related most similarly. The linear regression relating Tissue Transglutaminase IgA and caffeine consumption was 0.76, and the t-test had a p-value of 0.16. There is no significant difference in the mean volume of tissue antibodies in those who consume either above or below the recommended amount of caffeine. The findings presented in this work may be of value for nutritional health specialist.

References

[1] *Impacts of Gut Bacteria on Human Health and Diseases.* **Zhang, Yu-Jie, et al.** 4, Hong Kong : International Journal of Molecular Science, 2015, Vol. 16. doi.10.3390/ijms16047493.

[2] *Gut microbiota's effect on mental health: The gut-brain axis.* **Clapp, Megan, et al.** 4, Lubbock : Clinics and Practice, 2017, Vol. 7. 10.4081/cp.2017.987.

[3] **Harvard School of Public Health.** The Microbiome. *Harvard T.H. Chan School of Public Health.* [Online] 2021. [Cited: December 9, 2021.] https://www.hsph.harvard.edu/nutritionsource/microbiome/.

[4] *Role of Gut Microbiota in Nutrition and Health.* **Valdes, Ana M.** 361, Nottingham : BMJ, 2017. 10.1136/bmj.k2179.

[5] *The Role of the Microbiome in Insomnia, Circandian Disturbance and Depression.* **Li, Yuanyuan, et al.** 669, Guangdong : Frontiers in Psychiatry, 2018, Vol. 9. 10.3389/fpsyt.2018.00669.

[6] **Lee, Jang-Eun and Lee, Eunjung.** Impact of Drinking Alcohol on Gut Microbiota: Recent Perspectives on Ethanol and Alcoholic Beverage. *Current Opinion in Food Science.* [Online] Elsevier, October 14, 2020. [Cited: December 9, 2021.] https://www.sciencedirect.com/science/article/abs/pii/S2214799320300783.

[7] **Gonzalez, Sonia.** Nutrient. *Long-Term Coffee Consumption Is Associated with Fecal Microbial Composition in Humans.* [Online] DMPI, May 2020. [Cited: December 9, 2021.] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7282261.

[8] **Huang, Zuyi.** Chapter 2, General Introduction of R Language. *ChE 5332: Biochemical Data Analysis by Dr. Zuyi (Jacky) Huang.* Villanova : s.n., 2021.

[9] Chapter 6, Statistical Data Modeling. *ChE 5332: Biochemical Data Analysis, by Dr. Z Huang, Chapter 6.* Villanova : s.n., 2021.

[10] Chapter 3, Statistical Data Analysis. *ChE 5332: Biochemical Data Analysis by Dr. Zuyi (Jacky) Huang.* Villanova : s.n., 2021.

[11] **Harvard School of Public Health.** Added Sugar in the Diet. *Harvard T.H. Chan School of PUblic Health.* [Online] The Nutrition Source, 2021. [Cited: December 9, 2021.] https://www.hsph.harvard.edu/nutritionsource/carbohydrates/added-sugar-in-the-diet/.

[12] Caffeine. *Harvard T. H. Chan School of Public Health.* [Online] The Nutrition Source, 2021. [Cited: December 9, 2021.] https://www.hsph.harvard.edu/nutritionsource/caffeine/.