Harry Barrs

Marie Snipes

Applied Linear Algebra

10 May 2024

Diet Optimization of Fiber and Sugar

**Introduction**:

In America today, over one third of the population suffers from obesity. Linked with heart disease, diabetes, and more life threatening conditions, obesity is undoubtedly a formidable problem for our country. In order to address this issue, many have turned towards diet and the foods we consume to explain the widespread problem. In this project I will similarly focus on diet and finding and creating an optimal diet to help provide potential solutions for individuals struggling with obesity. According to Dr. Chiadi E. Ndumele, "elevated sugar consumption is an important contributor to weight gain." For this reason, I will attempt to create a diet with the idea of minimizing sugar as to help mitigate weight gain in an individual. In addition to sugar consumption being linked with obesity, recent studies have shown the importance of our microbiome, in our overall health as well as ability to correctly digest food. Although our microbiomes are a very under researched topic and are inherently difficult as they differ for every individual, certain organisms within individuals' microbiomes or rather the lack thereof, have been linked with obesity. In a talk given by Epidemiologist, Dr. Tim Spector, Spector explores the topic of our microbiomes, highlighting that although different for everyone, they all share a reliance on fiber. Though evidence and information on our microbiomes is still a relatively new area of research, I am highly interested in this topic. For this reason, in addition to exploring a diet with minimal sugar, I will also attempt to explore another diet in which we maximize fiber, as to help feed and grow one's microbiome.

In attempting to create diets with sugar and fiber in mind, I have decided to utilize a dataset which includes a vast number of food options as well as nutritional information. Taken directly from the online source, https://data.world/craigkelly/usda-national-nutrient-db, the data set I used was adapted from the USDA National Nutrient Database. Data Scientist, Kraig Kelly, cleaned and abbreviated the original database into an easier format. This allowed me to download a .csv file which I could manipulate using Microsoft Excel. Included in this dataset were more than 8000 food options as well as information on macronutrients, minerals, and vitamins pertaining to 100 g of the food.

Using the Simplex Method for linear optimization, I will create two diets. First I will minimize sugar and then I will maximize fiber. Due to the size and scope of the data, I will use MatLab for all of my calculations. In both scenarios, I will use the same set of scientifically based constraints pertaining to lower and upper limits for nutrient consumption. Though the constraints and system could be designed for any individual, I have focused on the demographic of adult males and more specifically, used myself as the case study for which I derived information for caloric boundaries. Although I initially intended to create fully usable and reasonably healthy diets, I was ultimately unable to mathematically find a solution within the constraints I originally intended. My final solution was arrived at by removing most of my upper bound constraints which allowed for extreme suggestions within the solution. That being said the results are neither uninteresting nor wholly worthless.

**Setup**:

In the following section, I will not only detail the structure of my project, but also the evolution and changes I had to make in order to arrive at my final solution. Firstly, after briefly looking at my dataset, I noticed a section of baby-foods within the 8000+ food options. Because I am focusing on creating a diet for an adult male, I decided to remove these options from the food list. Moving on to setting up our linear program(s), since we are working with a dataset of different food options and attempting to create diets, our variables x1, x2, x3,...,x7941 refers to the amounts of each different possible food option. Note that 7941 was my final number of food options and that in reality, my initial setup had a few hundred more variables which I later discarded. So, our objective functions will be:

$$\text{Sugar: } 0.06*x1 + 0.06*x2 + 0*x3 + \dots + 0*x7941 = f\_sugar$$

$$\text{Fiber: } -1*(0*x1 + 0*x2 + 0*x3 + \dots + 0*x7941) = f\_fiber$$

The coefficients in each equation are respectively the amount of grams of sugar or fiber per 100 grams of each food option (xi). Because MatLab's linprog function naturally minimizes, the entire objective function for fiber is multiplied by -1 in order to maximize fiber. In both functions, f_sugar and f_fiber will be the total grams of sugar or fiber.

As for my constraints, I will start with the calories and macronutrients. Because I am focusing on the demographic of adult males and using myself as the case study, I proceeded by calculating my caloric maintenance level. This refers to the amount of calories needed to maintain one's current weight. For adult males, one's caloric maintenance level can be calculated with the following equation:

$$\mathbf{((X/2.20462)*24)*(Y)}$$

X = Weight in lbs.

Y = Activity Factor

| CATEGORY | DESCRIPTION | ACTIVITY FACTOR |
|---|---|---|
| Sedentary | Little or no exercise | 1.2 |
| Lightly Active | Light exercise or sports 1-3 times per week | 1.375 |
| Moderately Active | Moderate exercise or sports 3-5 times per week | 1.55 |
| Very Active | Hard exercise or sports 6-7 days per week | 1.725 |
| Extremely Active | Very hard exercise, physical job or training 2x/day | 1.9 |

Note that this is not a perfectly exact calculation and that the activity factor is also up to the user to be honest. The previous chart and equations were sourced from lectures given by Kenyon College Professor, Emily Heithaus, in her class: Human and Sport Nutrition HSPS 171.00. Since I weigh 190 lbs and am Moderately Active (1.55), my caloric maintenance level is about 3206 kcal. Because I, myself, am not trying to lose weight, I modeled a relatively tight constraint for calories giving the program a leeway of 100 kcals on either side of my maintenance. As for macronutrients, these are the overarching nutrient types of protein, fat, and carbohydrate. For these constraints I utilized recommendations again sourced from Professor Emily Heithaus' class Human and Sport Nutrition HSPS 171.00. Given these percentages I calculated upper and lower gram amounts based on my upper and lower caloric bounds of 3306 kcal and 3106 kcal.

| Macronutrient | Percentage Range |
| --- | --- |
| Carbohydrate | 45-65 |
| Protein | 10-35 |
| Fat | 20-35 |

Given these percentages I calculated upper and lower gram amounts based on my upper and lower caloric bounds of 3306 kcal and 3106 kcal. In summary, for calories and macronutrients, I constructed the following constraints:

Calories (kcal): $3106 \leq C1x1 + C2x2 + C3x3 + \dots + C7941x7941 \leq 3306$

Carbohydrate (g): $371.925 \leq C1x1 + C2x2 + C3x3 + \dots + C7941x7941 \leq 537.225$

Protein (g): $82.65 \leq C1x1 + C2x2 + C3x3 + \dots + C7941x7941 \leq 289.275$

Fat (g): $73.46666667 \leq C1x1 + C2x2 + C3x3 + \dots + C7941x7941 \leq 128.5666667$

In each of these constraints, C1, C2, C3, … ,C7941 refers to the respective value of calories, carb, protein, fat, per 100 g of each respective food option xi. With our caloric and macronutrient constraints in place, we can move on to the rest of our constraints regarding minerals and vitamins.

Firstly, the data set provides information on the following nutrients:

Vitamin A

Vitamin B6

Vitamin B12

Vitamin C

Vitamin E

Folate

Niacin

Riboflavin

Thiamin

Calcium

Copper

Iron

Magnesium

Manganese

Phosphorus

Selenium

Zinc

In order to create constraints for these nutrients, I researched for the medically recommended upper bound as well as the daily recommended value to serve as the lower bound. In this way, we can focus on creating healthy diets as we are ensuring that the diet will give the necessary nutrients, but not too much as to be harmful. HarvardHealth.edu provided all the necessary research backed bounds to create my constraints. Utilizing both their suggested upper limits as well as recommended daily intakes as lower limits, I created constraints for almost all previously listed nutrients. Some nutrients such as Vitamin B12 and Riboflavin as well as others, did not have upper limits as no research has yet shown harmful effects to over consumption. With these constraints created and in place as well as all variables being set to greater than or equal to zero, $x_i \geq 0$, my initial linear program was ready.

Unfortunately, MatLab was unable to find a feasible solution within these constraints. Seeing as all my bounds were research based and decently broad, this was highly disappointing and could suggest potential flaws within the data. After heavy manipulation and trial and error, I removed all upper bound constraints aside from those pertaining to the macronutrients and calories. In this way, I hoped to still create a calorically feasible diet. After removing all micronutrient based upper bounds, MatLab was indeed able to find a solution. However, within this solution were anomalies due to the inclusion of beverages in the dataset. The most pertinent example being that when maximizing fiber the solution suggested approximately 148,725 servings of Powerade Zero. Looking closely at the data, this illuminated several problems. First being that the 148,725 servings could be explained by the fact that Powerade Zero contains zero calories. It does however, contain small amounts of Vitamin C and other nutrients. So, because the Powerade is zero calories, the program was able to add as much as needed to fulfill certain vitamin constraints while not adding calories. This enabled the program to give its focus almost entirely on maximizing fiber without any repercussions. Additionally, the 148,725 servings is extremely high and undesirable in terms of it being impossible to actually consume. Working from this, I decided to remove all beverages from the data set, as drinks are not food, as well as attempt to place an upper bound on each individual variable. Unfortunately, the upper bound of 2 servings per item, which intuitively seems reasonable with the large number of options the program can choose from, was again impossible, returning no feasible solution. Still, although I was unable to solve the potential problem of having too much of one food option, removing beverages did help and allowed me to set up for my final results. Here are the constraints which I used to achieve my final result:

All variables must be positive:

$x1, x2, x3, \ldots, x7941 >= 0$

Calories and Macronutrients:

Calories (kcal): $3106 <= C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 <= 3306$

Carbohydrate (g): $371.925 <= C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 <= 537.225$

Protein (g): $82.65 <= C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 <= 289.275$

Fat (g): $73.46666667 <= C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 <= 128.5666667$

Micronutrients:

Vitamin A (mcg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 900$

Vitamin B6 (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 1.3$

Vitamin B12: $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 2.4$

Vitamin C (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 90$

Vitamin E (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 15$

Folate (mcg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 400$

Niacin (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 16$

Riboflavin (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 1.3$

Thiamin (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 1.2$

Calcium (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 1000$

Copper (mcg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 900$

Iron (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 8$

Magnesium (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 420$

Manganese (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 2.3$

Phosphorus (mg) $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 700$

Selenium (mcg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 55$

Zinc (mg): $C1x1 + C2x2 + C3x3 + \ldots + C7941x7941 >= 11$

**Results**:

Maximization of Fiber:

f_fiber = 537.225 g

| Variable xi | Food Name | Value |
|---|---|---|
| 277 | Salt, table | 2.997769724742787e+04 |
| 318 | Oil, wheat germ | 0.071134605485821 |
| 354 | Oil, soybean lecithin | 0.499529947776030 |
| 659 | Turkey, liver, all classes, raw | 0.558290792220784 |
| 2331 | Pork, cured, ham, steak, boneless, extra lean, unheated | 1.253664749710856 |
| 3617 | Beef, short loin, porterhouse steak, separable lean only, trimmed to 1/8" fat, all grades, raw | 0.333866272369156 |

| | | |
|---|---|---|
| 4604 | Game meat, antelope, raw | 0.043075705892025 |
| 4678 | Veal, variety meats and by-products, thymus, raw | 0.728221629571751 |
| 7807 | Gums, seed gums (includes locust bean, guar) | 6.949870633893920 |

Minimization of Sugar:

f_sugar = 0 g

| Variable xi | Food Name | Value |
|---|---|---|
| 277 | Salt, table | 2.763816426090772e+04 |
| 414 | Margarine-like, vegetable oil spread, fat-free, tub | 11.396962195565628 |
| 1854 | Acerola, (west indian cherry), raw | 0.049662593834851 |
| 3311 | Epazote, raw | 1.857231272959591 |
| 5194 | Leavening agents, baking powder, double-acting, | 17.726473558850483 |

| | | |
|---|---|---|
| | straight phosphate | |
| 6839 | Beef, New Zealand, imported, liver, cooked, boiled | 12.263086054375437 |

**Result Analysis**:

Note that both the tables for the sugar and fiber optimization contain table salt. Table salt again contained zero calories in the data set. After seeing this result, I filtered for any remaining items with zero calories. There were four. I deleted these and attempted to run the MatLab code again. Without table salt, linprog could not find a feasible solution. So, I reverted back and the four remaining zero calorie foods in the dataset. However, table salt is the only one which showed up in our results. Again, the effect on the result of a food item with zero calories, is that it allowed the table salt to be in vast quantities and essentially remove my minimum constraints on a few of the minerals like iron and copper and magnesium.

**Fiber**:

Looking at our results we can see a pretty clear breakdown between fats, proteins, carbs. The oil fulfills the fat constraints, the meats fulfill the protein constraints and the seed gums fulfill the carb constraint. As for the maximization of fiber, this was literally exclusively met by the roughly 700 grams of seed gums which the program suggests eating. Looking at the different meats, I suspect that they each help fulfill my nutrient minimums. Because I ultimately was unable to use any of my initial micronutrient upper bounds, many of the micronutrients are blown out of healthy proportion into likely dangerous extremes. For example, the diet contains 4499.375 mcg of Vitamin A, which is about 1500 mcg over the suggested upper limit of consumption. This fact in tandem with the inclusion of zero calorie items (table salt) allowed for the maximum fiber value to be over 500 g.  This value is way too high for safe consumption. With this fact in mind, the diet can in now way be recommended at face value.  However, we can still learn from the results.  Seed gums are extremely fiber dense, around 75% by mass.  This seems to be a solid option for someone seeking to increase their fiber intake.

**Sugar**:

Moving on to sugar, we again have some interesting results.  Similar to the fiber optimization, we can see a clear fat source (margarine), protein (beef), and carb (leavening agents).  The other items seem to have been included to help fulfill remaining gaps in nutrient lower bounds.  Along with the table salt, the leavening agent is a strange inclusion.  This option, though not zero calories, is not really food.  It is moreso a tool in cooking.  The 17 plus servings of this item is wholly ridiculous.  Unfortunately, since I was unable to find a feasible solution with upper bounds on my individual variable amounts, the program allowed for such high unreasonable values.  Looking at the beef option this aligns with minimization with respect to sugar as meats tend to contain zero sugar.  All of the other options were also chosen clearly as options which contain zero sugar as our minimum value is zero.  This in itself is not a bad or unrealistic value as most sources suggest that males consume between 0 and 30 grams of added sugar per day.  That being said, with the leavening agents and the salt, we can again in no way recommend this diet, at face value, to anyone.  However looking at both tables we can see a consistency in that both chose a liver option.  The fiber optimization choosing turkey liver and the sugar optimization choosing beef liver.  Although an uncommon food choice  in the USA, looking at the data, both of these options are indeed nutrient dense and potentially good healthy food choices.  Though, not in vast quantities as suggested by the sugar optimization diet.

**Conclusion**:

Initially we set out to mathematically create usable diets both minimizing sugar consumption as well as maximizing fiber consumption in order to help combat obesity. In this pursuit, we failed as neither of the diets created are wholly usable. Without upper bound nutrient constraints and limitations on the amounts of each individual food item, the linear program was unable to create a realistic diet. Although we were able to come to a tangible result it was not as we intended. The results provided did give potential insights into intelligent food choices for those seeking to increase their fiber intake or decrease sugar intake while maintaining a healthy diet. Though I cannot be sure, the inability of the linprog function to find a feasible solution with our initial constraints suggest either a mistake in my inputs or errors in the raw data. Having checked, rechecked, and cross referenced the nutrient constraint inputs in the program as well as the available research; I am led to believe that there are flaws within the data set. Given more time, I would be highly interested in applying a similar method to a different set of data. Still, we did arrive at a result which we were able to discuss and analyze.

Works Cited


Centers for Disease Control and Prevention. (2021, November 28). *Get the facts: Added sugars*. Centers for Disease Control and Prevention. https://www.cdc.gov/nutrition/data-statistics/added-sugars.html#:~:text=Americans%20are%20eating%20and%20drinking,2%20diabetes%2C%20and%20heart%20disease.

Eppes, E., Krautheim, A. M., & Boone, K. (2019, October 3). *Obesity in the U.S.* Food Research & Action Center. https://frac.org/obesity-health/obesity-u-s-2#:~:text=The%20latest%20data%20indicate%20that,tend%20to%20increase%20with%20age.

Harvard Health Publishing. (2020, August 31). *Listing of vitamins*. Harvard Health. https://www.health.harvard.edu/staying-healthy/listing_of_vitamins

Heithaus, E. (2023a). *Basics for Metabolism*. Lecture, Gambier; Kenyon College.

Heithaus, E. (2023b). *Making More Nutritious Choices*. Lecture, Gambier; Kenyon College.

Ndumele, C. E. (2021, November 1). *Obesity, sugar and heart health*. Health. https://www.hopkinsmedicine.org/health/wellness-and-prevention/obesity-sugar-and-heart-health

Spector, T. (2019, February 27). *What role does our microbiome play in a healthy diet? - with Tim Spector*. YouTube. https://youtu.be/-LUuqxQSaFQ?si=jtCm3T5rXdzcX-zM

*USDA National Nutrient DB - dataset by Craigkelly*. data.world. (2016, December 20). https://data.world/craigkelly/usda-national-nutrient-db

WebMD. (2022, November 2). *Vitamins and minerals: How much should you take?*. WebMD. https://www.webmd.com/vitamins-and-supplements/vitamins-minerals-how-much-should-you-take