# MathWorks Math Modeling Challenge 2018


## Adlai E. Stevenson High School–

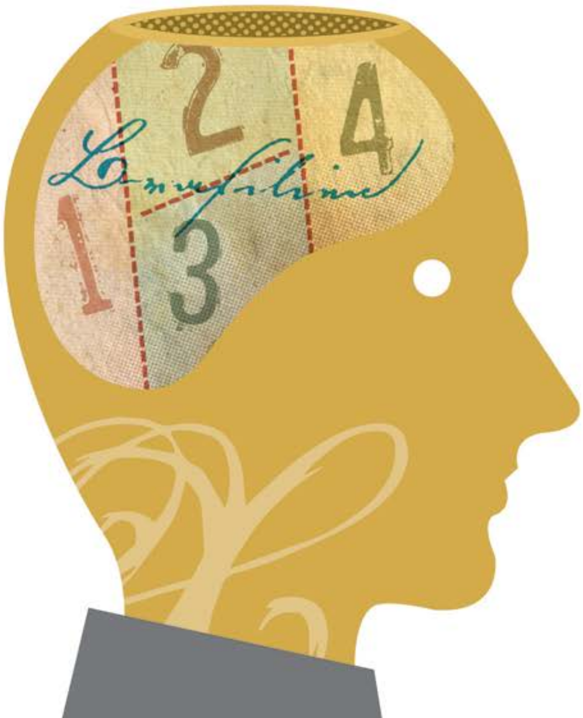
Team # 10265 Lincolnshire, Illinois

Coach: Paul Kim

Students: Albert Cao, Andrew Hwang, Deepak Moparthi, Joshua Yoon, Haoyang Yu


**MathWorks Math Modeling Challenge Finalist**

**$5,000 Team Prize**



# Foo(d) Fighters

Team Number: 10265

March 2018

# 1 Executive Summary

In 1798, Thomas Malthus predicted that food production would grow at a linear rate while population would grow exponentially. In modern times, we as a society have proved that wrong, but a growing problem that exists as a result is the increasing amount of food going to waste every year. Out of the food produced, only approximately 60% is actually eaten, with the rest being classified as food wastage.

We were initially tasked with developing a state based model to calculate if the food wasted could feed the group classified as food-insecure. Taking Texas as a test case, we found there was enough food wastage, monetarily speaking, to provide for all food-insecure individuals in one instance. Following this discovery, we developed a more complex model to see whether a sufficient amount of calories from each food group would be delivered to each person. We determined that there was a surplus in grain that could be supplied, but fruits/vegetables and proteins required deeper analysis. For fruits/vegetables, we determined that the excess corn would likely provide for the population, and for proteins, 0.8% of the food-insecure population would not receive the minimum of calories. However, we decided that the 52 calories a day that would not be guaranteed is small enough such that food-insecure people could acquire these calories through other means.

Following our work in the first part, we modeled the food wastage generated by a household given certain characteristics, which we deemed to be income bracket, age group of the oldest member in household, and size of household. We then created three intermediate indices, Income, Age Group, and Household Size, which independently predict the dollar food waste value of a household. Then, we trained a multivariate linear regression model to synthesize the intermediate indices altogether into a final model to predict the food wastage of four cases given to us in the prompt. Using our model, we found that a single parent with a toddler and a $20,500 income had a $242.19 yearly food wastage, and the family of four with a $135,000 income had $1043.56 yearly food wastage.

In part 3, we were asked to suggest possible solutions to the problem of food wastage in a certain community and perform cost/benefit analyses. We had three primary proposals: composting, share tables, and rescheduling gym times. When analyzing the effect of composting, we considered the discrepancies in pricing for commercial composting, landfill processing, and the potential reduction in greenhouse gas emissions, and realized this would lead to a net deficit. Moving on to the concept of share tables, which allow untouched food like whole fruits and milk to be traded out between students with the leftovers being transported to a nearby food pantry, we analyzed their projected growth in use over time as a result of increasing awareness about food wastage. Finally, when considering the effect of rescheduling gym to be prior to lunch, we found no obvious costs and only potential benefits.

Ultimately, food wastage is an imminent yet mitigable problem with far reaching consequences. We found that food wastage, when streamlined, has the potential to eradicate food insecurity, and investigated where it is greatest and where to focus on the most. Furthermore, we developed feasible solutions to reducing food wastage in our own community followed by multiple cost benefit analyses. Times certainly have changed since 1798, but now it's time for food wastage to follow suit, too.

# Contents

# 2 Part 1: Red Hot Chili Peppers

## 2.1 Restatement of the Problem

In part 1, we are asked to do the following:

- Create a model that would allow a state to calculate the wasted food generated within it.
- Determine if the wasted food will feed the state's food-insecure population.
- Apply our model to Texas

## 2.2 Assumptions

- All waste food generated within Texas can be transported to another area of Texas without the food going bad.

  Justification: The longest straight line distance within Texas is about 800 miles [1]. Therefore, the longest drive that a truck would have to travel to transport waste food is less than one day, and it is fair to assume that this is feasible to do.

- For our initial check, we will see whether the monetary value of the waste food passes a benchmark of approximately $2 billion.

  Justification: According to Feeding America, in 2015, Texas needed approximately $2 billion more dollars to feed its food insecure population [2]. This number has likely not changed significantly in the two years since 2015, and if it has, Texas (or other states using this model) may swap this benchmark for their own, updated one.

- The food waste produced by consumers (in the consumption step) is essentially unusable.

  Justification: The waste at the consumption stage "includes losses and waste at the household level" [3]. Most of this waste is likely half-eaten food, food scraps, or old food that are thrown away by households. It would be economically unfeasible to implement a program to collect and distribute these foods and also unsanitary and unethical to give this food waste to food-insecure households. For sake of simplicity, we will ignore the part of consumption waste that comes from businesses like restaurants.

- We classify foods distributed to food-insecure citizens as dairy, grains, fruits/vegetables, and proteins.

  Justification: These are the major food groups.

- Dairy will not be considered when breaking down caloric intake for each food group.

  Justification: We are given only one data point for dairy products in the state receipts [14]. Furthermore, food-insecure individuals can partially provide for themselves, so we can assume that they can provide for their own dairy needs if the state provides the rest.

- Each person receives some food daily, and the minimum standard that a state should attempt to achieve is 1000 calories in grains, and 500 each from fruits/vegetables and proteins.

  Justification: Based on our own breakdown and the NCBI Nutritional Values [13]. These numbers can be easily changed in our model, but we used these values as they appeared reasonable.

- For each food group, one item within it is given each day.

  Justification: Over time, people will accumulate food and store some in refrigerators. Thus, they will not necessarily be eating the three food items given to them daily; it is just a simple distribution model that is easier to model and apply in real life.

## 2.3 Developing the Model - Initial Check

In our model, we will first provide a model that serves as an initial check of whether a state can provide for its food-insecure citizens. To do so, we use the data given of the total sales of each food product from the state receipts and the data given of the percent food of each category wasted [14]. In our assumptions above, we justified that food wasted at the consumption step was not viable for use. Thus, our waste percentages are as follows:

| Food Group | Waste Percentages |
|---|---|
| Cereal | 11.25% |
| Roots and Tubers | 52% |
| Oil Seeds and Pulses | 18% |
| Fruits and Vegetables | 38% |
| Meat | 13.5% |
| Fish and Seafood | 27.5% |
| Milk | 5.7% |

To simplify our calculations, we represent the total sales of each food within a vector $\vec{v_f}$, and create a vector $\vec{w}$ which would contain the waste percentages of each food. Since the first 5 rows of the data given correspond to cereal, the first 5 elements of $\vec{w}$ would be the waste percentage for cereal: 11.25%. Similarly, $\vec{v_f}(i) = 0.52$ for $i = 6, 7$, the waste percentages for roots and tubers, and so on and so forth for the rest of the categories.

To find the total monetary value of the food waste generated, we take the following calculation: $\vec{v_f}^T \cdot \vec{w} = \$2,525,890,820$. This value passes our initial benchmark, 2 billion dollars, which was outlined in one of our assumptions. This implies that the dollar cost of waste food is enough to provide food for all food-insecure citizens in the population of Texas. Also, note that to generalize to other states, we only need to change the state receipts for each product by using data from that state.

## 2.4 Developing the Model - Caloric Intake per Food Group

### 2.4.1 Introduction

However, the previous initial check is a rough estimate that doesn't take into account all factors that determine whether a state can satisfactory feed all food-insecure citizens.

One factor that we will now consider is the different types of food that each person needs to sustain a healthy lifestyle.

To determine whether a state can also provide a balanced, healthy meal for its food-insecure citizens, we first categorize food into 4 major groups: grains, fruits and vegetables, proteins, and dairy. At any center that distributes food to those who need it, we assumed that one item from each group would be given daily, and ideally, it would contain enough calories to satisfy a person's daily requirement from that food group.

In our model, we will determine the ideal daily serving sizes for each food type within fruits/vegetables, proteins, and grains, such that every food-insecure citizen will receive a serving of an item from each group every day and also get enough calories to satisfy daily needs for each food group. As this is not always feasible, for the food groups where we find that a person may not receive enough calories daily, we will analyze deeper by looking at the distribution of the average amount of food each food-insecure citizen is projected to receive over an extended a long period of time.

### 2.4.2 Amount of Food Waste for Each Item

To begin, we must find the raw amount of food waste that can be given to those who need it. In the previous section, when performing the matrix multiplication to calculate the total monetary value of waste food generated, we can also find the monetary value of waste food for each type of food we are given data for by looking at the individual terms in the, that is $\vec{v_f}^T(i,1) \cdot \vec{w}(1,i)$. For example, the monetary value of wasted sorghum in Texas is $\$54,068,513 = \$480,609,000 \cdot 0.1125$. From data we collected on cost of each food type from the U.S. Department of Agriculture Economic Research Service, we were able to find the total amount of each food wasted in a standard unit of measurement for that type of food, e.g., bushels, pounds, cups. We then converted these into units that are easier to understand, such as number of oranges and quarter pounds of meat, which are commonly sold at supermarkets and given to those who can not fulfill their dietary needs without assistance.

Below is a table illustrating our conversions for some of the food groups we placed under fruits/vegetables. Note that we use Texas in our example, but for other states, the values in the second column will change.

| Fruits/vegetables | Value of Waste (TX) | Cost/Unit | Conversion factor |
|---|---|---|---|
| Corn | \$454,549,920 | \$3.30/bushel | 58 ears of corn/bushel |
| Watermelon | \$27,550,000 | \$0.21/ | 0.0625 half melons/cup |
| Onions | \$21,771,340 | \$0.41/cup | 1 onion/cup |
| Grapefruit | \$19,163,400 | \$0.85/cup | 1 grapefruit/cup |
| Cabbage | \$13,255,920 | \$0.25/cup | 0.125 heads of cabbage/cup |
| Oranges | \$11,521,220 | \$1.04/pound | 3.48 oranges/pound |

By multiplying the values in the second and fourth column, then dividing by the value in the third, we can find the total amount of waste the state has for each food item. For example, there are 7,989,059,200 ears of corn and 22,545,176 grapefruits the state can give per year. We perform similar calculations for every food item in each food group to determine a set of values $A = \{a_1, a_2, \ldots, a_n\}$, where $n$ is the number of items in the food group (6 for grains, 17 for fruits/vegetables, and 9 for proteins).

### 2.4.3 Ideal Daily Serving Sizes for Each Item

As stated before, for each food group, we hand out some serving size that we determine later in our model of one of its items to each food-insecure citizen daily. Also, a person only receives one randomly selected food item a day, with a greater likelihood of getting items that the state has more servings of. To calculate the ideal number of each item (such as ears of corn, onions, oranges, etc.) we should give in one daily serving, and we also need to find data on the number of calories per item and then create another set $K = \{k_1, k_2, \ldots, k_n\}$ where $k_i$ is equal to the calories per item $i$. For fruits and vegetables, $a_1 = 7,989,059,200$ and $k_1 = 455$ as each ear of corn has 455 calories [23]. Once we had both the raw amount of each food item, and their caloric value, we could then proceed to finding the ideal amount to give to food-insecure citizens.

We also created a set $S = \{s_1, s_2, \ldots, s_n\}$ where each $s_i$ is a positive integer indicating how many of each item we give to a citizen if that is their randomly designed food item that day, and the $s_i$ are what we seek to optimize.

As mentioned in our assumptions, our goal is to guarantee that every person receives some amount of food daily, and for each food group, reach some minimum standard in terms of caloric intake. For this to be true, the following equation must hold:

$$\sum_{i=1}^{n} \frac{a_i}{s_i} = 365 \cdot p,$$

where in this case $p = 4,320,050$, as that is the number of individuals that are food-insecure in Texas. This equation must be true because $a_i/s_i$ represents the number of daily serving sizes we give of item $i$ to a person. Furthermore, we wish for $s_i \cdot k_i$ to be as close as possible for all $i$, since this is the number of calories one receives if they are given item $i$ on some day, and regardless of which item in the food group one receives, one should have a similar daily caloric intake.

To calculate the ideal values of $s_i$ we used a computer program, which is included in appendix A. For each item, we input $a_i$ and $k_i$ into the program, along with reasonable bounds for $s_i$ as it would be unreasonable to give an abnormally large amount of a specific food like chickens or bread on one day since it would not be possible to consume it all.

## 2.5 Results

### 2.5.1 Grains and Fruits/Vegetables

For the grains, we discovered that the set $S$ should be $S = \{2, 2, 6, 2, 2, 8\}$, where the six items in the grains group are sorghum, wheat, rice, oats, rye, and potatoes.

As $K = \{644, 644, 206, 607, 644, 163\}$ for grains, Texas would be able to supply food-insecure citizens $1288, 1236, 1214,$ or $1304$ calories a day as these are all the values of $s_i \cdot k_i$. All of these values are above the minimum of 1000 calories from grains a day that we determined, so Texas does not have to be concerned with providing its food-insecure citizens with enough grains. However, for fruit/vegetables and proteins, citizens were not guaranteed to be get enough calories.

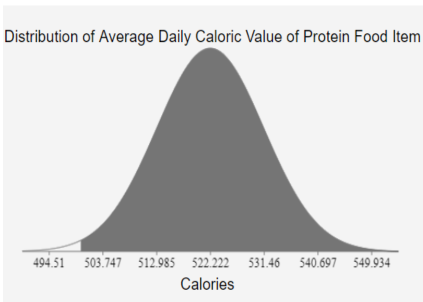
In the case of fruits and vegetables, we set a limit of one ear of corn daily per person because of its high caloric value. However, because of this restriction on the amount of corn, we noticed that there was an extreme amount of excess corn as we would not hand out all of it. This surplus of corn can be distributed alongside lower calorie items such as grapes and spinach so that caloric intake for fruits and vegetables can essentially be met daily. We believe that the factors responsible for the surplus of corn are that Texas is a major producer of corn; a fair amount of the corn is actually unusable corn waste, and because Texas likely exports some of the corn that it sells. In a more developed model, considering these factors would allow us to make a more accurate prediction on whether food-insecure citizens would receive enough calories from fruits and vegetables. As a result, we will not perform the additional analysis for the average amount of calories one would consume daily for sake of simplicity since we must also do it for proteins.

#### 2.5.2 Results - Proteins

For proteins, the program outputted an ideal solution where $S = \{2, 1, 1, 8, 2, 3, 2, 2, 1\}$ and the nine items are dry beans, cattle and calves, broiler chickens, chicken eggs, hogs, turkeys, farm chickens, catfish, and soybeans.
We also found that $K = \{240, 285, 210, 80, 275, 215, 270, 260, 830\}$ on the Food Composition Databases created by the U.S. Department of Agriculture. As a result, we get that each food-insecure citizen receives one of the following values in calories as their daily protein food item $T = \{s_1 \cdot k_1, s_2 \cdot k_2, \ldots, s_9 \cdot k_9\} = \{480, 285, 210, 640, 550, 645, 540, 520, 830\}$. We assume that each of these items will be received with equal probability for sake of simplicity and because the state has a very large amount of each item.
Note that the average of these elements is 522.222, which is greater than the minimum of 500 calories we set for proteins; there are multiple elements in the set $T$ that are less than 500, which means that dietary needs from the protein food group will not necessarily be satisfied daily.
Over the course of a long period of time, a person is randomly given one of the nine different items each day, which is equivalent to getting one of the caloric values in $T$ as his/her daily protein food item caloric intake. Over one year, we may consider the average caloric values of all items received in a year, as being the average of a random sample of 365 elements chosen from an extremely large multiset with equal amounts of each of $t_i$, which has $\mu = 522.222$ and $\sigma = 176.480$. Note that the sample is selected randomly, smaller than the size of the population, and sufficiently large, so we are able to apply the central limit theorem. Thus, we can conclude that the sampling distribution of the mean daily caloric values is normal and has $\mu_{365} = \mu = 522.222$ and $\sigma_{365} = \frac{\sigma}{\sqrt{n}} = 9.237$.

Note that as stated before, the minimum caloric intake that we aim to provide is 500 calories daily from the protein food group. Thus we wish to calculate the following:

$$N(\mu, \frac{\sigma}{\sqrt{n}})\, P(x \geq 500) = N(522.22, 9.237) P(x \geq 500).$$




As the z-score of 500 is

$$z = \frac{500 - \mu_{365}}{\sigma_{365}} = \frac{500 - 522.222}{9.237} = -2.406,$$

we find that

$$N(\mu, \frac{\sigma}{\sqrt{n}})\, P(x \geq 500) = P(z \geq -2.406) = 0.992.$$

This means that there is a 99.2% chance that one person receives above 500 calories daily on average over a year from protein food items. However, while we can expect approximately 99.2% of these people to receive the minimum standard, given that $p$ is over four million in Texas, this means that the probability that all food-insecure individuals receive an average of at least 500 calories daily on average over a year from proteins is essentially 0. However, note that the probability that entire population receives over 448 calories is essentially 1 as

$$(N(522.22, 9.237)P(x \geq 448.326))^{4320050} = (P(z \geq -8))^{4320050} = 1.$$

Note that since many food-insecure individuals are able to partially provide for themselves (in Texas 32% are above 185% poverty), we can safely conclude that for the expected 0.8% of the population that doesn't receive 500 average daily calories from proteins, the 52 daily calories that cannot be provided can be acquired through other means for most food-insecure people [25].

## 2.6 Conclusion

Based on our calculations of the probability that the entire food-insecure population receives the necessary food categories, we can safely conclude that the wasted food is likely enough to provide for everybody. Another factor that should be considered is the percentage of waste that is actually usable. For sake of simplicity, we assumed in our model that all waste in all stages except consumption were usable. This is unlikely to be completely true, so in a more nuanced model, we would analyze this factor more closely. Also, we would consider the factor of seasonal variation as some seasons may cause a surplus or shortage of season-specific foods. Some weaknesses of our model are the previously mentioned factors that we were unable to consider. On the other hand, our model's strengths include its simple initial check and a deep analysis on one important factor, the caloric intake for each food group, as it is important to provide all necessary nutrients as malnourishment also has harmful consequences. Another strength is that our model is easy to apply to other states since all that must be changed are the initial value of state receipts for each food item and $p$, the population of food-insecure individuals in the state. We would then replicate the process to determine if the food waste is truly enough to feed the food-insecure.

# 3 Part 2: Black Eyed Peas

## 3.1 Restatement of the Problem

In part 2, we are asked to do the following:

- Create a model that determines the amount of food waste a household generates in a year.
- Evaluate our model on the following households:
  - Single parent with a toddler, annual income of $20,500
  - Family of four (two parents, two teenage children), annual income of $135,000
  - Elderly couple, living on retirement, annual income of $55,000
  - Single 23-year-old, annual income of $45,000

## 3.2 Assumptions

- Psychographic characteristics are not considered in the model.

  Justification: We define psychographic characteristics as views toward food waste, such as those who value health over food and hence throw out expired products immediately, or those who know the cost of food wasted by America as a whole, and hence try to limit their waste as much as possible. While found to be important in determining the total amount of food wasted per household [5], due to its qualitative nature and limited data online, we forgo the factor for practical applications. For future directions, psycho-graphic characteristics should be considered.

- Food wastage will be measured in USD.

  Justification: For consistency measures, we aim to be able to compare the values of food wastage. While conceivable to measure the value of foods through their caloric value, the extremes of sugary products, for example, with high caloric content but little value to one's health, suggest that calories as a measure of comparison are too subject to confounding variables.

## 3.3 Developing the Model

In our model, we consider three factors: income bracket, age group of oldest member of household, and household size. According to Witzel and Hooge [4], these three factors are among the most influential factors that determine food wasted. We first model each of these factors separately, then combine them using multivariate linear regression to find the projected food wasted of each household as a function of these three factors.

### 3.3.1 Income Index

Within the modeling the effect of income on food wasted, we stratify food into four categories: Fruits/Vegetables, Grains, Dairy, and Protein. We split it this way because the amount of food purchased changes depending on income. For example, wealthier households tend to purchase more protein, while households with less than average income purchase more fruits and vegetables. With the provided data, we have the following table of income brackets and their respective annual expenditures on the four categories:

| Income Bracket | < \$15,000 | \$15,000 to \$29,999 | \$30,000 to \$39,999 | \$40,000 to \$49,999 | \$50,000 to \$69,999 |
|---|---|---|---|---|---|
| Grains | \$304 | \$374 | \$416 | \$521 | \$524 |
| Dairy | \$234 | \$298 | \$312 | \$384 | \$393 |
| Fruits/Vegetables | \$429 | \$549 | \$570 | \$704 | \$738 |
| Protein | \$617 | \$667 | \$690 | \$773 | \$840 |

| Income Bracket | \$70,000 to \$99,999 | \$100,000 to \$149,999 | \$150,000 to \$199,999 | \$200,000+ |
|---|---|---|---|---|
| Grains | \$611 | \$700 | \$828 | \$892 |
| Dairy | \$480 | \$565 | \$728 | \$666 |
| Fruits/Vegetables | \$919 | \$1,120 | \$1,345 | \$1,504 |
| Protein | \$1,012 | \$1,190 | \$1,465 | \$1,531 |

We then calculated the amount of food wasted per income bracket. The higher income a household has, the more likely they are to waste food because the wealthier you are, the more profligate you would be with food. To model this, we took countries from different parts of the world, ranging from the most developed countries to the least developed countries (i.e. U.S, China, Oceania, Latin America, Subsahara Africa, South Pacific Asia, etc), and used them to represent the various income brackets in our model. We did this because the income factor takes into account, ceteris peribus, only income, and the differences in cultures and ideals would be homogenized as a result. According to the New York Times, we see that this follows a somewhat linear regression [7]. In particular, in increasing order of income bracket, individuals in each income bracket throw away food at rates of:

$$4\%, 9\%, 11\%, 15\%, 20\%, 25\%, 31\%, 34\%, 39\%.$$

To combine these percentages with amount of each food group purchased by each income level, we normalized percentages by multiplying them by five, so that the \$50,000 to \$69,999 bracket represented the average of percent of food wasted per year. This makes sense, as the mean income in the United States is \$57,311 which falls in that range. From the provided data, we see that 27% of grains, 15% of dairy, 28% of fruits/vegetables, and 22% of protein are wasted. Hence, we get the following table for amount of food, in dollars, wasted for each income bracket. We define $Income_i$ as the Income Index, or the total food wastage in dollars per year per household as predicted for given income bracket $i$.

| Income Bracket (i) | $Income_i$ (In Dollars per Year per Household) |
|---|---|
| $< \$15,000$ | \$74.61 |
| $\$15,000$ to $\$29,999$ | \$200.76 |
| $\$30,000$ to $\$39,999$ | \$258.79 |
| $\$40,000$ to $\$49,999$ | \$424.09 |
| $\$50,000$ to $\$69,999$ | \$591.87 |
| $\$70,000$ to $\$99,999$ | \$896.16 |
| $\$100,000$ to $\$149,999$ | \$1316.18 |
| $\$150,000$ to $\$199,999$ | \$1753.82 |
| $\$200,000+$ | \$2142.43 |

### 3.3.2 Age Group of Oldest Member of Household Index

We first convert grams of food wasted to an equivalent dollar value using the NCBI's daily diet averages and a weighted average of proteins, lipids, and carbohydrates average prices [13]. Based on the dietary nutritional servings suggested by the NCBI, each day, on average, requires 50 grams of protein, 61 grams of lipids, and 300 grams of carbohydrates, with prices of \$0.505, \$0.73, and \$0.49 respectively per 100 grams. Using a weighted average by mass leads to a \$0.52 per 100 gram cost of food waste.

When modeling the effect of age group on food wasted, we consider three household age groups for the oldest person in the household: 18-34, 35-64, and 65+. According to a wastage study by Waste and Resources Action Programme, analysis of food wastage in UK households, households with oldest member aged 18-34 lead household waste with approximately 1900 grams per week, while 35-64 and 65+ waste 1450 g/wk and 1250 g/wk, respectively [9].

However, that is only the avoidable food wastage, which makes up 2/3 of the total food wastage in a household [8]. Scaling those three values to include total food wastage in a household and then associating the \$.0052 per gram conversion rate yields the following household wastage values for the age groups of oldest member aged 18-34, 35-64, and 65+. We define $Age_j$ as the Age Index, the total wastage in dollars per year per household predicted for a given age group of the oldest member in the household, $j$.

| Age Group of oldest member (j) | Avoidable Wastage (g/wk) | $Age_j$ (\$/yr/household) |
|---|---|---|
| $18-34$ | 1900 | \$770.64 |
| $35-64$ | 1450 | \$588.12 |
| $65+$ | 1250 | \$507.00 |

A sample calculation for the 18-34 group:

$$\frac{1900g}{wk \cdot household} \cdot \frac{52wk}{yr} \cdot \frac{\$.0052}{g} \cdot \frac{3}{2} = \frac{\$770.64}{yr \cdot household}.$$

### 3.3.3 Household Size Index

Household size has a critical impact on the amount of edible food disposed of per capita. Specifically, a negative correlation has been found between the amount of food waste per person and the number of individuals in a given household, i.e., despite the amount of household food waste growing as household size increases, the per person food waste decreases [5][6].

For example, in the same wastage study by Waste and Resources Action Programme, the following amounts of food waste were found [9][12].

| Household Size | Food Waste (g/wk/person) | % of given family size in America |
|---|---|---|
| 1 | 1500 | 28.13% |
| 2 | 2200 | 34.01% |
| 3 | 3300 | 15.44% |
| 4 | 4000 | 12.93% |
| 5 | 4600 | 6.00% |
| $\geq 6$ | 4800 | 3.51% |

Using the same pricings per 100 grams of food as established in section 3.3.2, the waste generated per year by a given household of any size and a sample calculation are given below. We define $Size_k$ as the size index, or the total wastage per year per household as predicted by the size of the household $k$.

$$Wastage = \frac{1500\text{g}}{wk \cdot household} \cdot \frac{\$0.52}{100g} \cdot \frac{52wk}{yr} = \frac{\$405.60}{yr \cdot household}.$$

| Household Size (k) | $Size_k$ |
|---|---|
| 1 | \$405.60 |
| 2 | \$594.88 |
| 3 | \$892.32 |
| 4 | \$1081.60 |
| 5 | \$1243.52 |
| $\geq 6$ | \$1297.92 |

## 3.4 Putting it all Together

Now confronted with three variables that affect the yearly food wastage of a given household, we proceed to train a multivariate linear regression model in order to input the indices of Income Bracket, Age Group of Oldest Member of Household, and Household Size as a function with a single output, the predicted waste of a given household. The data used for the regression includes the Income Bracket, Age Group of Oldest Member of Household, and Household Size based on a study of over 1000 participants on their yearly food waste [15].

The multivariate linear regression model has the form

$$W_t = b_0 + b_1 \cdot Income_i + b_2 \cdot Age_j + b_3 \cdot Size_k,$$

where $W_t$ is the predicted yearly wastage of a given household based on the $Income_i$, $Age_j$, $Size_k$, and $b_r$ for $r = 0, 1, 2, 3$ are coefficients of the intercept and each predictor's regression slope. Note that this model only requires the input of one's income, age, and household size.

In the above equation, we adjust the values of $b_i$ for $r = 0, 1, 2, 3$ to minimize mean squared error between the observed household waste from the data of the above survey and the wastage function, $W_t$.

Training our function with Excel, using 25 randomly selected data points from the survey for time sake, we find the following multivariate linear regression model,

$$W_t = -42.355 + 0.774 \cdot Income_i + 0.206195 \cdot Age_j - 0.05 \cdot Size_k.$$

## 3.5 Applying the Model

Now, we aim to solve the original problem and apply our model. To illustrate, let us evaluate our model on one of the households specified by the problem: a single parent, with a toddler, annual income of \$20, 500. Based on the given information, the single parent has an Age Index ($Age_{18-34}$) of \$770.64, Income Index ($Income_{\$15,000-\$29,999}$) of \$200.76, and a Size Index ($Size_2$) of \$594.88 (index values referred to are found on page 12 and 13).
Therefore, the predicted yearly food wastage this household would have would be

$$W_t = -42.355 + 0.774(\$200.76) + 0.206195(\$770.64) - 0.05(\$594.88) = \$242.19.$$

Applying this model to all the households, we find the following results:

| Household Description | Yearly food wastage ($W_t$) |
|---|---|
| Single parent, with a toddler, \$20, 500 income | \$242.19 |
| Family of four (two parents, two children), \$135,000 income | \$1043.56 |
| Elderly couple, retired, \$55,000 income | \$490.55 |
| Single 23-year-old, \$45,000 income | \$424.51 |

## 3.6 Validity of Models

In consideration of the final model, we find high levels of accuracy in this prediction. Despite the risks of using multivariate regression, such as potentially introducing (or failing to consider) variables that may or may not affect the total yearly household food waste, the high adjusted $R^2$ value of 0.887 on real individuals confirms that 88.7% of all variability in the actual yearly household food waste per household can be accounted for by Income Bracket, Age Group of Oldest Member of Family, and Household Size.

Further, in assessing the coefficients of each factor, we find that household income and age affect one's waste greater than household size (which has a negative coefficient.) This makes sense because of the limiting factor of one's income determining how much food one can buy, thus making it a larger factor. And naturally, one's age is then next important as the eating habits and dietary needs of an individual largely shape the foods one would buy and the quantity. Lastly, the negative value of the coefficient of household size reflects the findings of [5] which show the return of increasing household size results in a decreasing addition of additional household waste.

For future directions, we hope to continue training our multivariate regression using more data points to strengthen our model. While not in accurate as is, greater accuracy can be established following the use of all 1000 data points. For the sake of time, we only train with 25 points.

To continue, we can further establish the validity of intermediate indices. Due to the nature of each index defined as the dollar amount of food waste generated by any given household, we can sum each of the Age Indices, each of the Income bracket indices, and each of the Household size indices, and all three sums should equal one another as they each calculate the total net food waste of households across the country. As we will show now, the similarity between each of these total costs further corroborate our model.

### 3.6.1 Total Food Wastage Based on Income Bracket

Using the income bracket model, we calculate the nationwide food wastage in dollars by combining the income distribution with waste contributed. According to the U.S Census Data, the income bracket follows the approximate distribution [18]:

| Income Bracket (i) | Percent Distribution of US Households |
|---|---|
| $< \$15,000$ | 11.2% |
| $\$15,000$ to $\$29,999$ | 9.6% |
| $\$30,000$ to $\$39,999$ | 9.4% |
| $\$40,000$ to $\$49,999$ | 12.9% |
| $\$50,000$ to $\$69,999$ | 17.0% |
| $\$70,000$ to $\$99,999$ | 12.3% |
| $\$100,000$ to $\$149,999$ | 14.1% |
| $\$150,000$ to $\$199,999$ | 6.6% |
| $\$200,000+$ | 7.0% |

We then perform a weighted average to find the total amount of food wasted in the United States. For each income bracket, we multiply the food wasted, $Income_i$, by the percent distribution, to get average food wasted per household, $W_i$. We arrive at

$$W_{\text{income}} = \sum \text{i} \cdot \% \text{ Distribution} = \$768.8122,$$

as the average food wasted per household. Multiplying by the total number of households in the United States, 126, 244, 000 we get

$$768.8122 \cdot 124,244,000 = \$97,057,927,376.80,$$

or about $97 billion worth of wasted food that the United States outputs.

### 3.6.2 Total Food Wastage Based on Household Size

Using a similar approach as the prior section for Income based modelling, we refer back to section 3.3.3:

| Household Size | $Size_k$ | % of given family size in America |
|---|---|---|
| 1 | \$405.60 | 28.13% |
| 2 | \$594.88 | 34.01% |
| 3 | \$892.32 | 15.44% |
| 4 | \$1081.60 | 12.93% |
| 5 | \$1243.52 | 6.00% |
| $\geq 6$ | \$1297.92 | 3.51% |

Using a weighted average based on the total number of households in the U.S., 126, 244, 000, the sum total of wasted food in the United States is

$$\sum_{k=1}^{k\geq 6} Size_k \cdot (\text{\% of a given family size in America}) \cdot 126,244,000 = \$90,175,033,550.$$

#### 3.6.3 Total Food Wastage Based on Age Group of Oldest Member of House

Now using the age group model, we calculate the wastage in dollars per person per year using the baselines of \$770.64, \$588.12, and \$507 per household per year. Dividing these by the average number of people in an American household (2.58), followed by a weighted average based on the age distribution breakdown of oldest member of household age group (28%, 52%, and 20% respectively) and multiplying by American population (322,179,605) yields 58.5 billion [22]. Thus, we arrive at $\$77,797,833,100$ as the total nationwide food wastage.

### 3.7 Conclusion

To summarize, our final model, $W_t = -42.355 + 0.774 \cdot Income_i + 0.206195 \cdot Age_j - 0.05 \cdot Size_k$, to predict the household food wastage found the following estimates for the specified households.

| Household Description | Yearly food wastage ($W_t$) |
|---|---|
| Single parent, with a toddler, \$20, 500 income | \$242.19 |
| Family of four (two parents, two children), \$135,000 income | \$1043.56 |
| Elderly couple, retired, \$55,000 income | \$490.55 |
| Single 23-year-old, \$45,000 income | \$424.51 |

Further the three nationwide total food wastage calculations based on each independent factor in section 3.6 are \$97 billion, \$90 billion, and \$78 billion, respectively. The relative closeness of each value, only being off at most from the middle value by 13% affirm the validity of each of our modelled indices as the calculation of the net wastage done independently arrive at roughly the same value. Variance in these net costs can be attributed to the fact that our individual models for each index are discrete and thus offer room for potential over/under calculation within each sum.

And in conclusion, our model's high $R^2$ values when compared to real households, intermediate checks on individual indices, and deep analysis on three broad ranging factors lend credibility and the potential to extend this model easily given sufficient information and time to further train our multivariate linear regression to be more accurate.

# 4 Part 3: Smashing Pumpkins

## 4.1 Restatement of the problem

Taking our school to be our community, we are tasked with the following:

- Model potential solutions to food waste in our community.
- Perform a cost-benefit analysis.

## 4.2 Assumptions and Justifications

- Our high school is our community.

  Justification: Since the prompt involved choosing a community, we chose our high school because of our mutual familiarity with it.

- Energy recovered from the landfill once wastage has been sent there is negligible.

  Justification: There are only 86 energy recovery facilities in 25 states across the country, and each facility takes $100 million to construct; as a result, it is impractical to rely on them to recover a nontrivial amount of energy [16].

## 4.3 Background Information

In order to understand the possible impacts of food wastage and lack thereof, we must first understand the cycle of producing and decomposition of food. The production of food that will ultimately go to the landfill leaves a carbon footprint of 3.3 billion carbon dioxide equivalents, which includes the process of growing the crops, processing, transportation, etc. 30-40% of food goes to waste. This food mostly ends up in landfills where it is given an anaerobic environment in a span of 6-12 months, after which it decomposes to form methane, another greenhouse gas. We see the emergence of two distinct criteria to consider when investigating food wastage: monetary costs/benefits and environmental impacts.

## 4.4 Developing Potential Solutions

### 4.4.1 Composting

Composting is a means by which wastage that otherwise would end up in a landfill is taken to a commercial processing facility and repurposed as fertilizer, therefore cutting down on subsequent carbon emissions. However, on average, the fees necessary for a composting company to take one 96-gallon tote of wastage are almost double what it takes for transportation to a landfill, $18 as opposed to $34. Extended to several more totes expended by the school, this monetary gap only widens. At the same time, however, the carbon emissions are significantly lower. A 96-gallon tote of food waste would hold 364.8 lbs of waste [23], and considering landfills lead to 5.1 million tons of methane, or 107.8 carbon dioxide equivalent tons, a tote going to the landfill would yield .296 tons of carbon dioxide equivalent [20]. A 2015 price report projects a medium case price of $20 for one ton of carbon dioxide based on its environmental impacts [21], so one tote of food wastage would cost \$5.91 environmentally if sent to the landfill, which would be avoided if composted.

Ultimately, after a cost benefit analysis of commercial composting, using one tote of food wastage as a basis, we see a \$12 deficit per tote as a result of composting.

#### 4.4.2 Share Tables

Share tables have been used in many schools to reduce the amount of food thrown away by decreasing the initial amount of unwanted food. This works by having a "Sharing Table" located near the cafeteria where students can place unopened food and drinks that they choose not to eat or drink. This table provides an opportunity for other students to take additional, cost-free helpings of food or beverages.

There are little to no costs associated with this strategy. No advertising costs would be necessary because, as a high school community, the presence of the share tables itself is enough to increase awareness the food waste problem. A supervisor would be not be needed since the food is free to the students, and students would be able to volunteer to bring unused food to local pantries and charities.

These minimal costs are clearly outweighed by their obvious benefits. According to [14], approximately 40% of food is wasted at schools, and up to 40% is a result of taste preferences. As the Natural Resources Defense Council estimates, even a 15% reduction in food losses can result in an additional 25 million Americans fed [24]. While this statistic is nationwide, we see the same impact in our relatively small local community - the implementation of share tables can lead to

$$2.5 \cdot 10^7 \text{people} \cdot \frac{40\%}{15\%} \cdot \frac{74,936 \text{ local citizens}}{326,018,263 \text{ US citizens}} = 15,323$$

more people fed around the our high school community,

#### 4.4.3 Increasing Participation

A vital part to the effectiveness of share tables is the participation of the students on the program. Since not all students will be eager to take part initially, we model participation as a function of time in order to see the actual benefits that will come about in the future. This relationship, linked to awareness, can be likened to the affect of advertisement on the sale of a product, where the advertisements is the share table itself. The participation starts out relatively low, and the more advertisement is displayed, the more people become involved until it reaches a certain cap. This cap will never change because there are certain people who bring lunch from home, people who are adverse to share tables, or other people who have extenuating circumstances. This pattern leads us to the conclusion that participation can be effectively modeled by a four parameter logistic regression.

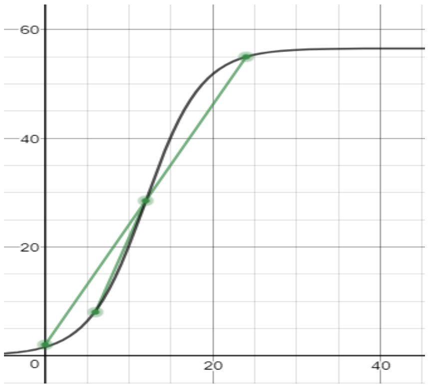
In order to create an exact representation, we would need four data points from the implementation of the program into the high school itself. Because of the lack of data, we can create reasonable estimates based on our school's participation in charity drives and collective initiatives and use the following points:

| Time (Months) | Participation |
|---|---|
| 0 | 2% |
| 6 | 8% |
| 12 | 28.5% |
| 24 | 55% |

Thus, our model is:

$$P = \frac{56.53}{1 + 35.25 \cdot e^{-0.30 \cdot t}}$$

where $P$ is participation and $t$ is the time in months, as shown in the figure on the right.




### 4.4.4 Rescheduling Gym Times

Another strategy to reduce food waste seen mostly in elementary schools is moving lunch to after recess. While, of course, high school students do not have recess, they are required by law to take gym. Hence, we propose rescheduling gym periods for everyone to be prior to lunch time, in order to increase consumption of food during lunch and thus reducing food wasted due to wrong planning of meals, which 20% of the population surveyed in [11] ticked as a reason for wasting food.

Just as with share tables, this option does not have any costs, yet there are again potential benefits. While we do not have any specific data about rescheduling gym times for high school students and its effect on food waste, we predict an increase in food consumed, and hence a reduction in food wasted.

## 4.5 Conclusion

In overview, we covered three strategies that we thought might serve to effectively manage food waste in our high school: composting, share tables, and rescheduling gym times. Our calculations for the cost-benefit analysis of composting led us to the result that the operational costs outweighed the environmental benefits, in particular, a $12 net cost per tote. In contrast, share tables and rescheduling gym times did not have any noticeable costs of implementation or maintenance, and thus we considered the long term benefits of increasing participation and the final impact of the program.

# 5 References

[**1**] “Environment.” Texas Almanac. Texas State History Association, n.d. Web. 4 Mar. 2018.
[**2**] “Map the Meal Gap.” Feeding America. N.p., n.d. Web. 04 Mar. 2018.
[**3**] “Global Food Losses and Food Waste.” Extent, Causes, and Prevention. Food and Agriculture Organization of the United Nations, n.d. Web. 4 Mar. 2018.
[**4**] Aschemann-Witzel, Jessica. “Consumer-related Food Waste.” Sustainability. N.p., n.d. Web. 4 Mar. 2018.
[**5**] “Volume to Weight Conversion Factors.” US EPA. Office of Resource Conservation and Recovery, n.d. Web. 4 Mar. 2018.
[**6**] Chakona, Gamuchirai, and Charlie M. Shackleton. “Local Setting Influences the Quantity of Household Food Waste in Mid-sized South African Towns.” PLOS ONE. Public Library of Science, n.d. Web. 04 Mar. 2018.
[**7**] Sengupta, Somini. “How Much Food Do We Waste? Probably More Than You Think.” The New York Times. The New York Times, 12 Dec. 2017. Web. 04 Mar. 2018.
[**8**] “Household Food and Drink Waste in the UK.” WRAP. Waste and Resources Action Programme, n.d. Web. 4 Mar. 2018.
[**9**] “Household Food and Drink Waste: People Focus.” WRAP. Waste and Resources Action Programme, n.d. Web. 4 Mar. 2018.
[**10**] “Why Americans Waste so Much Food.” Ohio State News. N.p., 21 July 2016. Web. 04 Mar. 2018.
[**11**] Jörissen, Juliane, Carmen Priefer, and Klaus-Rainer Bräutigam. “Results of a Survey among Employees of Two European Research Centers in Italy and Germany.” MDPI. Multidisciplinary Digital Publishing Institute, 05 Mar. 2015. Web. 04 Mar. 2018.
[**12**] “Distribution of U.S. Households by Size 1970-2016 — Statistic.” Statista. N.p., n.d. Web. 04 Mar. 2018.
[**13**] “Table 2.” NCBI. American Journal of Clinical Nutrition, n.d. Web. 4 Mar. 2018.
[**14**] “Cash Receipts by State.” Economic Research Service. USDA, n.d. Web. 4 Mar. 2018.
[**15**] Neff, Roni A., Marie L. Spiker, and Patricia L. Truant. “Wasted Food: U.S. Consumers’ Reported Awareness, Attitudes, and Behaviors.” PLOS ONE. Public Library of Science, n.d. Web. 04 Mar. 2018.
[**16**] “Energy Recovery from the Combustion of Municipal Solid Waste (MSW).” EPA. Environmental Protection Agency, 14 Apr. 2017. Web. 04 Mar. 2018.
[**17**] Harvey, Chelsea. “The Enormous Carbon Footprint of Food That We Never Even Eat.” The Washington Post. WP Company, 28 Mar. 2016. Web. 04 Mar. 2018.
[**18**] Jessica L. Semega, Kayla R. Fontenot, and Melissa A. Kollar. “Data.” Income and Poverty in the United States: 2016. N.p., 01 Sept. 2017. Web. 04 Mar. 2018.
[**19**] “Measuring the Benefits of Composting Source Separated by Organics. CM Consulting, n.d. Web. 4 Mar. 2018.
[**20**] “Carbon Emissions from Waste Measured in EPA Greenhouse Gas Inventory.” Recycling Waste Management World. N.p., 22 Oct. 2015. Web. 04 Mar. 2018.
[**21**] “Carbon Dioxide Price Report 2015.” Synapse Energy. N.p., n.d. Web. 4 Mar. 2018.
[**22**] “Population Distribution by Age.” The Henry J. Kaiser Family Foundation. N.p., 22 Sept. 2017. Web. 04 Mar. 2018.
[**23**] Food Composition Databases Show Foods List. N.p., n.d. Web. 04 Mar. 2018.
[**24**] Gunders, Dana. “Wasted: How America Is Losing up to 40% of Its Food from Fark to Fork to Landfill.” NRDC Issue Paper, n.d. Web. 4 Mar. 2018.
[**25**] “Overall Food Insecurity in Texas by County in 2015.” Map the Meal Gap. Feeding America, n.d. Web. 4 Mar. 2018.

# 6 Appendix A

Listing 1: C++ code used in Part 1 to find the optimal serving size of each food item in Texas

```
#include <iostream>
#include <fstream>

using namespace std;

long long maxPer[20], b[20], calPer[20];
long long amtPer[20], totalPer[20], store[20];
int n;
long long pop, min, max;

void solve(int x)
{
        if (x == -1)
        {
                long long ans = 0;
                for (int i = 0; i < n; i++)
                {
                        ans += totalPer[i] / amtPer[i];
                }
                if (ans >= 0)
                {
                        int cnt = 0;
                        min = LLONG_MAX;
                        max = 0;
                        for (int i = 0; i < n; i++)
                        {
                                store[i] = amtPer[i] * calPer[i];
                                if (store[i] > max && maxPer[i] > 1)
                                    max = store[i];
                                if (store[i] < min && maxPer[i] > 1)
                                    min = store[i];
                                cnt++;
                        }
                        if (max - min < 500)
                        {
                                for (int i = 0; i < n; i++)
                                {
                                        cout << amtPer[i] << endl;
                                        cout << store[i] << endl;
                                }
                                cout << max - min << endl;
                                cout << ans - 365 * pop << endl;
                        }
                }
```

```
                return;
        }
        for (int i = maxPer[x]; i >= 2; i--)
        {
                amtPer[x] = i;
                x--;
                solve(x);
                x++;
        }
}

int main()
{
        ifstream fin("moodys.txt");
        fin >> n >> pop;
        for (int i = 0; i < n; i++) // max for each
        {
                fin >> maxPer[i];
        }
        for (int i = 0; i < n; i++) // cal for each
        {
                fin >> calPer[i];
        }
        for (int i = 0; i < n; i++) // numerator
        {
                fin >> totalPer[i];
        }
        solve(n-1);
        system("pause");
}
```