CS 4460 Homework 1

David Thomson

January 18th, 2017

Initial Questions

- Which manufacturer has the highest variation in nutritional value among its cereals? Which manufacturer has the least variation?
- Which cereal is the healthiest overall per serving?
- Which manufacturer produces cereals with the highest amount of fiber?
- What does a value of -1 mean? Is it different depending on the attribute?
- Is the shelf number arbitrary, or is there a correlation between the shelf number and some other attribute of the cereal?

Insights

- For all 9 nutritional attributes (calories, protein, fat, sodium, fiber, carbohydrates, sugars, potassium, and vitamins) the average amount per cup across all cereals was greater than the average amount per serving. Note that this is true despite the fact that some cereals had a value of -1 for cups, meaning that some of the nutritional values calculated per cup were large negative numbers. If I knew the meaning of the negative numbers and corrected for them, it appears that the difference between nutrition per serving and nutrition per cup would be even greater. This leads the the question of why this difference occurs. Is it a statistical coincidence, marketing, economics, etc.?
- The healthiest overall cereal (see Process for definition of healthiest) is Fruit & Fibre Dates, Walnuts, and Oats from Post.
- Playing around with the weights of different nutrients (see Process) did not change the rankings as much as I though it would, leading to the question of what is the best way to evaluate the health of a cereal? Is there a general set of weights that will work for most purposes?
- Out of all the nutrients, sodium content had the highest variation across manufacturers, while other attributes such as sugar and calories had very little variation in comparison.
- Every manufacturer with multiple cereals has at least one cereal on every shelf except for Ralston Purina, which does not have any cereals on shelf 2.

1

Process

I began by taking the plain averages of the 9 nutrient columns. Then I created 9 new columns and used a simple Excel formula to populate them with the total amount of nutrient per cup for each nutrient. I assumed the original values were per serving. I took the average of these new columns and compared them to the original averages. Next, I tried to calculate the healthiest overall cereal. I used Excel's `RANK` function on each nutrient column according to the following preferences: low calories, high protein, low fat, low sodium, high fiber, low carbohydrates, low sugars, high potassium, and high vitamins. For each cereal I calculated the sum of the ranks in each column. Finally, I ranked these totals to determine that Post's Fruit & Fibre Dates, Walnuts, and Oats is the healthiest cereal per serving. I then decided to use a weighted sum of the rankings of each nutrient. For example, in one calculation I gave sugar content a weight of 4 and potassium a weight of 1, since it is generally a good idea to avoid excess sugar, and potassium needs can be met easily with bananas or avocados. Next I decided to calculate the average amount of each nutrient per serving per manufacturer. Then I looked at the differences between the manufacturers. Finally, I looked at the shelf numbers of the cereals and which manufacturers appear on which shelves to try and see a pattern.

Difficulties

The biggest challenge turned out to be a lack of creativity in dealing with the raw data. The information mostly consists of amounts of different nutrients, so I had trouble thinking of insightful calculations without the aid of visualization tools. The values of -1 were also problematic. In many cases I felt like they didn't affect the calculations very much since I tried to get a general overview. Most of the cereals had normal values, but it would have been nice to have a way to correct for the negative entries. I also found that my Excel skills could use improvement. Constantly copying around data, typing formulas, etc., without being able to make a simple graph has given me a greater appreciation for the power of visualization.