Statistics Project (Part I)
Total points for part I: 100

For this project, you will work in groups of 3-5 (you must work in groups) on an assigned topic. There is no required length to the project, simply complete the assignment to the best of your ability. Feel free to discuss your progress with me or submit work early for feedback.

Your group must choose a product from a local grocery store. For the product type you pick you must collect the following information:

a. Store you collected the information from (they can all be from the same store)
b. Serving per container
c. Serving size in grams(note the will probably not be the same for all products)
d. Cost per container
e. Type (For example: Campbells' Chicken noodle, Post Raison Bran, Breyers vanilla….)
f. Sodium or Sugar per serving
g. Fat per serving
h. Calories per serving

(for parts f, g, and h, you may need to collect a different kind of information depending on the product you pick, check with me before you collect your data if you are not sure what to collect)

BEFORE you collect any data you must decide how you want to compare your products. For example you may choose to compare products by:

- Compare one brand of cereal to another brand.
- Compare organic brands to non organic
- Compare the shelf locations (top shelve brands versus bottom shelve brands)
- Compare diet to non-diet.
- "Healthy" versus "Non-Healthy"

However you decide to compare your product you must collect at least 20 products for each of the two groups you have decided to compare. (So if you decide to compare cereal brands you will need information from 20 cereals of each brand, so a <u>total of 40 cereals</u>)

It is important to decide how you want to compare your products before you go collect your data. You may not have enough data if you do not make this decision before hand

All projects should include:

1. Paragraph 1: Your paper should begin with an introduction. This should explain what you are doing and why you are doing it, what you hoped to learn or explore with the project. (5 points)

2. Paragraph 2: A summary of how the data was collected (was it random or a version of random). Be clear how you data was collected, indicate your plan, how you implemented the plan and any issues with collecting your data. You will most likely run into problems collecting the data, tell me any problems. If random, you need to be specific, not just say it was random. Did you use a random number generator,…..Systematic or Stratified is probably the way to go for most projects. (5 points)

3. Paragraph 3: A table of all of your data which will include the mean, median, mode, standard deviation, 5 number summary and the upper and lower fences for each of the numeric values you collected (the values in parts f, g, h). NOTE: if the servings sizes are different you must adjust for the different sizes, you need to make all the values for calories/fat/sugar (the data in parts f,g and h) the same units (so convert everything to a per gram value, you can do this by dividing the calorie column by the serving size column that was recorded in grams) (30 points)

4. Paragraph 4: Box-plots, you will need to do two box plots for each of the variables in parts f, g and h (so a total of 6). The idea is to compare these values based on how you split the data set. (So does one brand have similar or different sugar content to another). You will have 3 sets of 2 boxplots, draw each set on the same axes. Mention the shape of the data sets based on the box plots. (18 points)

5. Paragraph 5: A discussion of what you learned. I am looking for an explanation of the data and what it tells you. Do not tell me how to find a mean or standard deviation, tell me what a mean or standard deviation tells you. Discuss any variations in the data set, any outliers. I am looking for a discussion of what you found out and why the data supports this. Your discussion should include any other information you think the reader needs to understand the data.
    - If you use the word "healthier" you should indicate what you mean by such a term, be specific.
    - Use numbers to support your claims and talk about the mean and standard deviation (a good time to use the range rule, Chebyshev's or empirical rule).
    - Mention any outliers or unusual values and what their effect on the rest of the data may be.
    - Discuss what the boxplots tell you. You must discuss all three variables from parts f, g and h. (40 points)

6. Paragraph 6: A list of the group members and what each did on the project. (2 points)

Except for the graphs, I expect the papers to be typed, neat and easy for anyone to read and understand.