ALY-6050   Week 1 Project

1. This Week 1 Project has 100 points.
2. Solutions should consist of an Excel workbook and a Word document. Please attach both files when submitting the project.
3. Perform all calculations and analysis In the Excel workbook.
4. In the Word document, write a report summarizing the results obtained in your Excel workbook.
5. The project will be graded according to the following rubric:

The submission of each weekly project will consist of an Excel workbook and a Word document (an R script file is optional)– a minimum of two submissions that have been submitted as attachments. For each weekly project, students should complete their analytic work in an Excel workbook, and write a minimum of 1000 words in a Word document describing their findings. The Word document should consist of a title page (including student's name, assignment title, course number and title, the current academic term, instructor's name, and the assignment completion date), and a reference page. The Word submission of each project will consist of three sections:

(i)   Introduction
(ii)   Analysis
(iii)   Conclusion

The weekly projects will be graded by using the following criteria:

(a) Excel Workbook (50 points)
  (i)   Problem set-up and modeling (50%)
  (ii)   Problem solution and accuracy (50%)
(b) Word Document (50 points)
  (i)   Description of the problem introduction (20%)
  (ii)   Description of the problem analysis (30%)
  (iii)   Description of conclusion (30%)
  (iv)   Writing Mechanics, Title page, and References (20%)

Project:

The quality analyst of a candy manufacturing company randomly selected 103 bags of the company's 47.9 grams (advertised) candy bags to perform a statistical analysis. Each bag contains about 50 to 65 candies that are of six different colors: red, orange, yellow, green, blue, and brown.

The data has been saved in the Excel workbook ALY-6050 – Week 1 Project – Data.xlsx.

1. Perform numerical descriptive statistics of the weights, and create a frequency histogram along with a cumulative relative frequency line plot of the total weight in each bag. What probability distribution do you anticipate to best fit the distribution of weights? Justify your answer statistically.
2. Perform numerical descriptive statistics of the number of candies in these bags, and create a frequency histogram along with a cumulative relative frequency line plot of the total number of candies in the bags. What probability distribution do you anticipate to best fit the distribution of the total number of candies in these bags? Justify your answer statistically.
3. Create Boxplots of the number of candies in a bag for each of the six colors. Among the six boxplots, which color seems to be more unusual than the other colors? Justify your response statistically.
4. Perform numerical descriptive statistics of the percentage of defective items in each bag, and create a frequency histogram along with a cumulative relative frequency line plot of the percentage of defective items in each bag. What probability distribution do you anticipate to best fit the distribution of the percentage of defective items in each bag? Justify your answer statistically.
5. The manufacturer has advertised that each bag contains 47.9 grams of candy. Does there exist sufficient evidence from this sample to accept the manufacturer's claim?
6. An important quantity in quality control is the value of the standard deviation. For example, considering the amount (weight) of candies in each bag, a large standard deviation indicates that the fluctuations in the overfilling or underfilling of candy bags are significant. The company wishes to make its best efforts to minimize both overfilling (which, results in a loss of revenue) and underfilling (which, results in a loss of reputation). In your judgement, does there exist sufficient evidence to conclude that the standard deviation of the weights is more than 0.5 grams? Justify your answer statistically.
7. From the given data and considering only the number of defective items in each bag, select 20 random samples with each sample consisting of 30 bags. Calculate the mean and the variance of the number of defective items (per bag) for each sample. Create a frequency histogram of the sample means and a frequency histogram of the sample variances. In your opinion, what specific probability distribution best fits the distribution of the sample means, and what specific probability distribution best fits the distribution of the sample variances? Justify your answers statistically.