**Problem Set 6**

This problem set covers material from classes 15 and 16, including material from OIS sections 6.2, 7.1, and 7.2. It makes use similar data to that we discussed in class. Partial credit may be given for answers that are correct in part, but not in full. This problem set is due on Gradescope by Wednesday October 20, 2021 at 11:59 PM.

**Part I: Rasmussen vs. Rasmussen 2 (40 pts.)**

In addition to the approval rating numbers from over the summer for President Trump that we used in class, both Rasmussens came out with polls right before the election.

These poll results are summarized below:

* Rasmussen Reports: Biden 48%, Trump 47%, 1500 likely voters.
* RMG Research (Scott Rasmussen’s firm): Biden 51, Trump 44% 1,200 likely voters.

1. For this question, we are first going to focus on **Joe Biden’s support** in these two polls.
   1. Please calculate and interpret the 95% confidence interval for the difference in **Biden’s support** in the two polls. Please show your work. Does there appear to be a significant difference between the two polls? How do you know? (For this problem, you can assume that the necessary conditions have been met.) (10 points.)
   2. Now please calculate the z-score for the difference in **Biden’s** support in these polls and then find and interpret the associated p-value using the cutoff at the 95% confidence level. Please show your work for the z-score and explain how you got the p-value. (15 points.)
   3. If we were using a one-tailed less-than hypothesis test at the 95% confidence level (i.e., that Biden’s support was less in the Rasmussen Poll than in the RMG poll), would we be able to reject the null hypothesis? What if we were using a one-tailed less than hypothesis test at the 90% confidence level? What do you risk by reducing your confidence level? Please find p-values to support your answers and discuss how you arrived at these values. (15 pts.)

**Part II: A fruiT-Test (35 pts.)**

Halloween is just around the corner. In that spirit, the *FiveThirtyEight* candy\_rankings package has [data](https://fivethirtyeight.com/videos/the-ultimate-halloween-candy-power-ranking/) on a variety of attributes of candies and their popularity. Among these attributes is whether the candy contains fruit.

1. Among the 38 candy bars that contain fruit, the mean winning percentage (i.e., its popularity) is 44.1 and the standard deviation is 10.3. (20 pts.)
   1. Please find and interpret a 95% confidence interval for the mean popularity of candy bars with fruit. Please explain how you found your t-critical value. Does this t-distribution closely approximate a normal distribution? Why? (10 pts.)
   2. Let’s say that your null hypothesis was that the average popularity score of candy bars with fruit was equal to 50. Find a t-score and p-value at the 95% confidence level to test the alternative hypothesis that the mean rating of candy bars with fruit is different from 50. Please show your work, interpret your p-value, and explain how you found the p-value. (10 pts.)
2. FiveThirtyEight also classified whether candies had chocolate or not. Among the 37 candies with chocolate, the mean winning score was 60.9 and the standard deviation was 12.8. Is there evidence that the mean score for candies with chocolate is greater than 60? Please show your work, interpret your p-values and explain how you found the p-value (15 pts.).

**Part III: A “Peared” T-Test (25 pts.)**

1. The World Apple and Pear Foundation has collected data on the number of pears (and apples) produced for each country from 2003 to 2013. The United States and Italy tend to be the second and third largest producers of pears. (25 pts.)
   1. The mean difference in the number of pears produced (i.e., US pears produced minus Italy pears produced) is -25,541 and the standard deviation of the difference is 88,069.41 for the 11 years included in the data. Please conduct and interpret a “peared” t-test at the 95% confidence level to examine whether there is a difference between the number of pears produced in the United States and Italy. Please show your work and discuss how you found your answer (15 pts.).
   2. The below histogram shows the distribution of the difference between the number of pears produced in Italy and the number of pears produced in the United States. Given the number of years included in the study, should this histogram give us pause in conducting this test and interpreting our results? Why or why not? Please provide evidence to support your conclusion (10 pts.).

