**Problem Set 3**

This problem set covers material from the two classes after the exam, including material from OIS Sections 5.3 and 6.1; this problem set also draws upon information from OIS Section 5.2. 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.

**Part I: Hypothesis Testing Framework (20 pts.)**

1. You are a researcher who would like to examine the following scenarios. What would be the hypothesis testing framework that you would use? (12 pt.)
   1. You wonder if people who live in rural areas differ in their support of single payer health insurance compared to those who live outside rural areas. (4 pts.)
   2. You are interested in testing whether voters are less likely to vote when there is rain on Election Day. (4 pts.)
   3. You think that civil wars might be more likely to take place when a country experiences a severe economic depression. (4 pts.)
2. Having collected data, you would now like to evaluate the evidence. (8 pts.)
   1. Let’s say that the data you collect is really different from your null hypothesis for parts 1a and 1b. What would you do? What would you do if you found evidence that was fairly similar to the null expectation in part 1c? (4 pts.)
   2. What happens if the null hypothesis was actually true for part 1a, but you rejected it? What if the alternative was actually true for part b, but you failed to reject? Which of these potential errors is more serious, if either? (4 pts.)

**Part II: Confidence Intervals and Hypothesis Testing (70 pts.)**

1. On Sunday July 12, the [University of Texas at Tyler](https://projects.fivethirtyeight.com/polls/20200712_TX.pdf) released a poll of the upcoming presidential and Senate elections in Texas. First, I would like you to consider the presidential results. (40 pts.)
   1. In this poll, among likely voters (sample size= 1,677) Joe Biden had 48% of the vote, while Donald Trump had 43%. Find and interpret a 95% confidence interval for Biden’s percentage and discuss whether you have evidence to suggest that Biden’s percentage differs from 50% and how you know this. (12 pts.)
   2. Now, please find the z-score for a 95% confidence interval associated with Biden’s percentage here. (12 pts.)
   3. After finding the z-score, what are two options for finding the p-value associated with this estimate? Using one of these, find and interpret the p-value. Can we reject the null hypothesis here? (8 pts.)
   4. Using a one-sided hypothesis test associated with the same z-score, do you have evidence to suggest with 95% confidence that Biden’s percentage in the poll is *less than* 50%? How do you know? (8 pts.)
2. Now, I would like you to consider the Senate results in the poll. In the same sample of 1,677 likely voters, Republican incumbent Senator John Cornyn led Democrat M.J. Hegar by a 42 to 29% margin. (The Texas Senate primary is on Tuesday, I used Hegar here because she led State Senator Royce West in a separate primary survey in this poll.) (30 pts.)
   1. Can we conclude, with 95% confidence that John Cornyn’s percentage in the poll *differs* from 50%? What is the p-value associated with this estimate? Please show your work or explain how you got your answers along the way. (20 pts.)
   2. The university decides that they would like a margin of error of no larger than 1.5 percentage points for this poll. Is their sample size big enough? How big would they have to make it in order to have a margin of error smaller than this number? Please show your work, but double check using a calculator (or R). (10 pts.)

**Part III: P-Values and Confidence Levels (10 pts.)**

1. Is there any reason why we have to have a 95% confidence level when hypothesis testing? What does Denworth suggest as an alternative to this to convey surprisingness? What happens, conceptually, if you increase your confidence level from 95% to 99%? (10 pts.)