**Problem Set 4**

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. You should be able to do everything but question 4b after the class on September 23/24. We will discuss this material early next week.

**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 cities differ in their support of public transit 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 snow on Election Day. (4 pts.)
   3. You think that civil wars might be more likely to take place when a country experiences a severe drought. (4 pts.)
2. Having collected data, you would now like to evaluate the evidence.
   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 Tuesday September 22, the [Atlanta Journal-Constitution](https://www.ajc.com/politics/ajc-poll-race-for-president-senate-contests-in-georgia-too-close-to-call/FCIZO2M5ZNH2XAQBRMAXJLFQVE/) released a poll of the upcoming presidential and Senate elections in Georgia. First, I would like you to consider the presidential results. (40 pts.)
   1. In this poll, among likely voters (sample size= 1,150) Joe Biden and Donald Trump both had 47% of the vote. Find and interpret a 95% confidence interval for either candidate’s percentage and discuss whether you have evidence to suggest that their 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.)

First, you need to calculate the standard error so we need n. Here, n is 1,150.

* 1. 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.)
  2. 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.)

1. Now, I would like you to consider the Senate results in the poll. In the same sample of 1,150 likely voters, Republican incumbent David Perdue led Democrat Jon Ossoff 47% to 45%, with Libertarian Shane Hazel at 4%. Under Georgia law however, if no candidate receives 50% in November, the top 2 candidates will advance to a January runoff. (30 pts.)
   1. Given the importance of the 50% threshold, Ossoff’s campaign is interested in whether we can discern whether their candidate’s support is discernibly below this threshold. Can we conclude, with 95% confidence that Jon Ossoff’s percentage in the poll *is less than* 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 AJC decides that they would like a margin of error of no larger than 2.5 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.)