Instructions:

* Please upload your answers to Blackboard in Word and/or GraphPad documents by midnight on Friday, March 18.
* You may use any available resources (class notes, videos, textbooks, etc). However, you may not work together or get advice from other people. The work you submit must be your own.
* Show your work. If you use the right approach but make a careless error, we can give you partial credit. If you have the right idea, but use the wrong number from a previous part of the question, you will get partial credit.
* Each part of each question is worth 2 points.

**Problem 1:**

Researchers are investigating the distribution of HbA1c in a population. In a sample of 100 people, they find that sample mean HbA1c is found to be 5.2% and the sample standard deviation (SD) of HbA1c is found to be 0.5%.

1. Provide an interpretation of the estimated SD (what does it tell you about the distribution of HbA1c in this population?)
2. Calculate the standard error (SE) of the estimated mean?
3. Provide an interpretation of the SE? (what does it tell you?)
4. Construct a 95% confidence interval for the mean HbA1c
5. Interpret the interval you constructed in part D.

**Problem 2:**

A study is being planned to estimate the mean change (MC) of a biomarker after exposure.

The null hypothesis is that: H0: MC=0.

Let stand for the estimated mean change based on a sample.

Suppose based on statistical theory, it is known that under the null hypothesis and given the sample size, the distribution of is normal with mean equal to 0 and SD= 4.

1. `What is the probability that will be between -4 and +4?
2. What is the probability that will be greater than 6?

The study was conducted and was found to equal to 9.

1. Based on this finding, what is the p-value for the null hypothesis that MC=0?
2. Interpret the p-value you obtained for part C (if you did not obtain a p-value, make one up for the purposes of this question).

**Problem 3.**

A research team wanted to investigate whether relaxation training helps reduce anxiety in individuals experiencing stress. In a pilot study, they recruited 30 people at random from a group of 100 who have "high stress" jobs. The 30 people were divided into two groups. One group of 15 was used as the control, receiving no training. The second group of 15 received the relaxation training. After 6 months, the subjects in each group were evaluated based on an anxiety inventory. The resulting data can be found on the excel file called Problem3data.xlsx. Using Graphpad or other software, answer the following questions:

1. What is the sample mean and SD of the anxiety score in each group.
2. Provide a graph which illustrates the findings.
3. State properly the null and the alternative hypotheses of this study
4. Calculate a p-value for the null hypothesis. Explain the approach you used.
5. Interpret the p-value you calculated in Part D

**Problem 4.**

Consider the results of the pilot study described in Problem 3.

1. If the team wants to have 80% chance to detect the same difference at 5% level of significance between each group as that in the pilot study, how many participants are needed for each group?
2. The team has been advised that the clinical application of their method may tolerate the chance making type-I error no more than 1%. If the rest of the conditions remain to be the same as above, what is the minimal number of participants needed for each group?
3. Under the same conditions as question B above, what would be their chance to detect a difference as few as 2 between the two groups?
4. Suppose that the control group has recruited 5 fewer participants than that of the relaxation group. To maintain the statistical power at 80%, how many more participants does the research team need for the relaxation group?