CS 3130 / ECE 3530: Probability and Statistics for Engineers **Due Tu 11/21**

Homework 7: Estimation and Confidence Intervals

Instructions: Submit a single R Markdown file (.Rmd) of your work on Canvas by 11:59pm on the due date. You may also submit diagrams, drawings, etc. as image files (.png, .jpg, .gif)—they must be formatted into your .Rmd document (we won't look at them separately). **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

You may discuss the concepts with your classmates, but write up the answers entirely on your own. Do not look at another student's answers, and do not show your answers to anyone.

1. Exercise 19.4, pg. 295. **Only part (b) is required!**
   Part (a) will be 20 pts extra credit (must show a full explanation).

2. Your friend flips a fair coin n times and tells you how many heads showed up. She does not tell you how many times she flipped the coin. She repeats this experiment 10 times (each experiment has n flips), and reports to you the number of heads for each experiment: $x_1, x_2, \ldots, x_{10}$.

   (a) Give an unbiased statistic, $\hat{n}$, that you would use to estimate n. Include an argument for why the bias is zero. **Hint:** Start with the mean statistic $\bar{X}_{10}$. What is its expected value, and how would you modify this statistic it to make it unbiased?

   (b) Simulate this experiment 10,000 times with $n = 25$ (each simulation will produce a list of 10 numbers). Use your statistic is part (a) to estimate n for each simulation (you should end up with 10,000 values for $\hat{n}$). **Hint:** See the `EstimationExamples.Rmd` code. This is similar to the simulation we did of the mean of a Gaussian.

   (c) Plot a boxplot of your 10,000 $\hat{n}$ values. Draw a horizontal red line where the true value of n is. Are your $\hat{n}$ values centered roughly around the true value of n?

   (d) Plot of histogram of your 10,000 $\hat{n}$ values. Using the `lines` command, plot the pdf of a Gaussian on top of your histogram, with μ and σ equal to the sample mean and standard deviation of your $\hat{n}$ values.

3. Exercise 23.3, pg. 356. **Calculate the confidence interval in R using the formula in the notes. You should use the `qnorm` function.**

4. Exercise 23.10, pg. 360. **Do your calculations in R. You'll want to use the `qt` function.**

5. In this problem you are going to analyze the built-in R data set `iris`. First, extract the sepal width of the virginica species and save it to a vector `x` using this command:

   ```
   x = iris$Sepal.Width[iris$Species == "virginica"]
   ```

   (a) Using a Gaussian approximation, what is the 95% confidence interval for the mean of `x`?

   (b) Using a Student t distribution, what is the 95% confidence interval for the mean of `x`?

   (c) Now assume that you only have the first 10 measurements. That is, create the vector `y = x[1:10]`. Repeat parts (a) and (b) for the vector `y`.

   (d) How did the decrease in sample size affect the results? Were the two different confidence intervals affected differently?