Project 3 **Due Friday, February 10**
Inference for Quantitative Variables **Math 222**

1. In 1986, researchers in Germany conducted a field study to explore whether driver characteristics are related to an aggressive response. The study was conducted at a busy intersection in Munich, West Germany, on two afternoons (Sunday and Monday) in 1986. The experimenters sat in a Volkswaggen Jetta (the blocking car) and did not accelerate after the traffic light turned green, and timed how long before the blocked car driver reacted (either by honking or flashing headlights). The response time (in seconds) is our variable of interest. Some values were censored in that the researcher stopped timing before the driver actually honked. This can happen if there is a time limit to the observation period and success has not been observed within that time period. The results are contained in the file Honking.txt.
   (a) Explore the distribution of the results of this data. Include some charts of the data, and explain your findings.
   (b) Which is larger, the mean or the median response time? Why is it larger?
   (c) Make a 90% t-distribution confidence interval for the mean response time. Explain clearly what population this confidence interval might apply to.
   (d) Are the conditions for using a t-distribution confidence interval satisfied by this data? Explain why or why not.

2. Unlike means, there are no formulas for a confidence interval for the population median response time in the example above. We can however use bootstrapping to try to say something about any population parameter, including a median.
   (a) Using R to re-sample, construct a bootstrap distribution for the sample median of studies like this one (10,000 re-samples should be enough, make sure to re-sample with replacement). Be sure to include a visual display of the shape of the distribution.
   (b) Which outcomes would be at the 5th and 95th percentiles on this bootstrap distribution? What is the 90% bootstrap confidence interval for the median?
   (c) Now use bootstrapping to find a 90% confidence interval for the mean.
   (d) How do the confidence intervals for the mean in this problem compare with those in the last problem?
   (e) How do the two confidence intervals for the mean compare with the one for the median? Is that the result you would expect?

3. In a 1993 paper, researchers studied a sample of people who claimed to have had an intense experience with an unidentified flying object (UFO). One of the many variables they considered was the IQ of the subjects. Suppose you want to test whether or not the average IQ of those who have had such a UFO experience is higher than 100, so you want to test:
   - $H_0 : \mu = 100$, versus

- $H_A : \mu > 100$.

The sample mean of the 25 people in the study was 101.6 with a standard deviation of 8.9. The resulting t-value is only 0.899 which is not statistically significant (the p-value is 18.9%).

Should the researchers have been surprised that their results weren't significant? What if they wanted to be able to detect a difference of 5 IQ points. Compute the statistical power of the test that the researchers carried out. Explain each step along the way. Be sure to clearly describe both the null model for the t-values and the alternative model. Recall that the alternative model will have a **non-central t-distribution** with non-centrality parameter:

$$\delta = \frac{\mu_A - \mu_0}{\sigma/\sqrt{N}}.$$

You can use the sample data to approximate σ in this formula.