Math 222 - Project 2 **Due Friday, February 2**

1. A teacher doubted whether his students could tell the difference between Coke and Pepsi, so he arranged an experiment. Each of his 21 students got three cups. Two cups contained one brand of cola, and the third cup contained the other brand. Which cup contained which brand was randomly determined for each student. Each student was asked to identify which cup contained the cola that was different from the other two. It turned out that 12 of the students successfully identified the odd cola.

   (a) Is this strong evidence that students really do better than just guessing? Back up your answer with an explanation of the results of the appropriate statistical test. Include a clear statement of hypotheses and the p-value calculation. Explain the procedure you used to get your p-value.

   (b) Calculate a 95% confidence interval based on these sample data. Clearly define the parameter that this interval estimates, and interpret the interval.

   (c) Describe what Type I and Type II errors mean in this situation.

   (d) Calculate the power of the teacher's test if the reality is that half of all students in the population can tell the difference.

   (e) If the experiment involved a sample of 100 students instead, then we could use a normal approximation. Calculate the power of this test at detecting a specific alternative hypothesis that half of students can correctly identify the difference when the sample size is 100.

2. 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. Save a copy of the file in a folder where your R-markdown file is located, and then enter the command:

```
honkingData = read.csv("honking.txt")
```

   (a) Explore the distribution of the results of this data. Include some charts of the data, and explain your findings. Be sure to describe the shape (skew, normality, etc.) of the data, and point out any outliers. You might also want to speculate about why different people may have reacted differently to the blocking car.

   (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.