Project 2 **Math 222**
Inference about Proportions **Due Friday, February 3**

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. The file SuperBowl.csv contains data about the winner and scores from every year since Super Bowl I. Save a copy of the file in a folder where your R file is located, and then enter the command:

```
superbowl = read.csv("SuperBowl.csv")
```

   to work with the data.
   (a) Make a histogram and a boxplot for the differences in scores between the winning team and the losing team for each year. Briefly give a summary of the shape, center and other features of this distribution.
   (b) Are there any outliers? What years, if any, were outliers?
   (c) Does this data look roughly normal? Make a normal quantile plot and comment on what you see.
   (d) Even though the data is clearly not normal, make a 95% prediction interval for the difference between the scores of the winning and losing teams in future superbowls. Explain what your interval means in words. *Warning: Unlike confidence intervals, prediction intervals are not robust against departures from normality, so take the results of this interval with a large grain of salt.*