Project 3 **Due Friday, February 9**
Inference for Quantitative Variables **Math 222**

1. 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 (the *margin of victory*. 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.*

2. 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
   - $H_A : \mu \neq 100$.

   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 38.8%).

   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.