STA9750 – Homework 1 [Due 09/26]

Exercise 1: *A group of scientists recorded the following measurements:*

1 4 2 3 6 9 1 3 9 3

Read in the data with SAS and

   i)   *Summarize the data: create a plot and find the mean and standard deviation.*
   ii)   *Find a 95% (normal) confidence interval for the mean. Interpret the results.*
   iii)   *The scientists are interested in knowing whether the mean is 5 or not. Is the mean significantly different than 5 at the 0.01 level?*

Exercise 2: *A pharmaceutical is interested in knowing whether their new treatment is significantly different than the current gold standard. They collected a sample of 40 individuals: 20 of them were assigned the new treatment, and 20 of them were assigned the current treatment. The outcome is on an ordinal scale that goes from 0 to 100, where 0 is "bad" and 100 is "great"*

Read in the dataset "pharma.csv" into SAS and

   i)   *Summarize the data: create a plot and find the means and standard deviations of the groups.*
   ii)   *Is the new treatment significantly different than the current gold standard?*
   iii)   *Is the t-test appropriate? Why or why not?*
   iv)   *Should the new treatment replace the current standard? Explain why or why not.*

Exercise 3: *A summer school is interested in knowing whether their program "works." They recorded scores in a math test before and after the summer school on 130 students.*

Read in the dataset "school.csv" into SAS and

   i)   *Summarize the data: create a plot and find the mean and standard deviation of the difference "after-before."*
   ii)   *Provide a 95% confidence interval on the difference "after-before."*
   iii)   *Would you recommend going to this summer school? Why or why not?*

Exercise 4: *[Source: openintro.org] A 2010 survey asked 827 randomly sampled registered voters in California "Do you support? Or do you oppose? Drilling for oil and natural gas off the Coast of California? Or do you not know enough to say?" Below is the distribution of responses, separated based on whether or not the respondent graduated from college. The data can be found in "drill.csv."*

   i) *Create a contingency table and a plot that summarizes the data.*

   ii) *What percent of college graduates and what percent of the non-college graduates in this sample do not know enough to have an opinion on drilling for oil and natural gas off the Coast of California?*

   iii) *Conduct a hypothesis test to determine if the data provide strong evidence that the proportion of college graduates who do not have an opinion on this issue is different than that of non-college graduates.*

Exercise 5: At some university, an introductory math course is taught by 3 different instructors. The math department is interested in knowing if there are significant differences in scores in the different sections. You can find the data on the course website under the name "scores2.csv". Plot and summarize the data and help the math department make a decision.