BIA 654 Homework 3

1. In an NYT/CBS poll, 56% of 2,000 randomly selected voters in New York City said they would vote for the incumbent in a certain two-person race.

   (a) Calculate a 95% confidence interval for the population proportion π. (Check whether the necessary 'assumption' is met.)

   (b) Carefully interpret the meaning of the confidence interval obtained in (a).

   (c) What is the margin of error?

   (d) Assume we had no prior knowledge about the true proportion π. We want to construct a 95% confidence interval for π with margin of error 2%. How large a sample is needed? How does the sample size change if we want to be 99% confident?

2. Many companies are experimenting with "flex-time," allowing employees to choose their schedules within broad limits set by management. Among other things, flex-time is supposed to reduce absenteeism. One firm knows that in the past few years, employees have averaged 6.3 days off from work (apart from vacations). This year, the firm introduces flex-time. Management chooses a simple random sample of 100 employees to follow in detail, and at the end of the year, these employees average 5.5 days off from work, and the sample standard deviation (SD) is 2.9 days.

   (a) Did absenteeism really go down, or is this just chance variation? Formulate the null and alternative hypotheses and carry out the testing.

   (b) Repeat the above for a sample average of 5.9 days and an SD of 2.9 days.

3. Open up the data file Therm.dat (or Therm.xls) uploaded in Canvas. Carry out the hypothesis testing to check whether the temperature measurement difference > 0 at significance level 0.01. (Don't forget to check whether an 'underlying assumption' holds or not, e.g., normal probability plot and Goodness-of-Fit Test for normality checking.)

4. (True, False) To make a t-test with 5 measurements, use Student's t-distribution with 5 degrees of freedom.

   Technical notes. The notion of "degrees of freedom" actually has to do with the notion of dimensions in linear algebra context (e.g., projection, subspace). Here's the idea behind the phrase. The standard error (SE) for the average depends on the standard deviation (SD) of the measurements, and that in turn depends on the deviations ($x_i - \bar{x}$, $i = 1, \ldots, n$) from the average. But the sum of the deviations has to be 0 (by the definition of mean $\bar{x} = (x_1 + \cdots + x_n)/n$), so they cannot all vary freely. The constraint that the sum equals 0 eliminates one degree of freedom. For example, with 5 measurements, the sum of the 5 deviations is 0. If you know 4 of them, you can compute the 5th–so there are only 4 degrees of freedom.