Homework 7: Estimation and Confidence Intervals

Instructions: Your answers are due at the beginning of class on the due date. You can either turn in a paper copy, or a pdf version through canvas. We highly recommend using latex (`http://www.cs.utah.edu/~jeffp/teaching/latex/`) for producing the assignment answers. If the answers are too hard to read (e.g. **pdf's from your phone's camera must be very clear and careful**) you will lose points (entire questions may be given 0).

Please make sure your name appears at the top of the page.

You may discuss the concepts with your classmates, but write up the answers entirely on your own. **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

1. Say a Utah pollster conducted $m = 16$ polls among people who voted in the 2012 presidential elections, and reports that 55% of the respondents would vote for Mia Love. But the pollster did *not* report how many people were interviewed in each poll, but always used the same number n. However, they did report how many people polled voted for Romney in 2012: $x_1, x_2, \ldots, x_{16}$. And say you know exactly 3/4 of all voters voted for Romney in 2012.

    (a) Give an unbiased statistic that you would use to estimate n. Please include a verification that the bias is zero.

    (b) What is the variance of this statistic?

2. In this problem you are going to analyze the built-in R data set `iris`. First, extract the sepal length of the setosa species and save it to a vector `x` using this command:

    ```
    x = iris$Sepal.Length[iris$Species == "setosa"]
    ```

    Answer the following: **Write down the R code you use to compute parts (a), (b), (c) – it should be very short.**

    (a) Using a Gaussian approximation, what is the 99% confidence interval for the mean of `x`? (Assume you know the true standard deviation of x is $\sigma = 0.3525$).

    (b) Using a Student t distribution, what is the 99% confidence interval for the mean of `x`?

    (c) Now assume that you only have the first 15 measurements. That is, create the vector `y = x[1:15]`. Repeat parts (a) and (b) for the vector `y`.

    (d) How did the decrease in sample size affect the results? Were the two different confidence intervals affected differently?