**Week 7: Hypothesis Testing with CLT**

1. All the State of the Union speeches, organized by year, are listed at the following address: <https://en.wikisource.org/wiki/Portal:State_of_the_Union_Speeches_by_United_States_Presidents>

The average word length in the English language is 5.1 words, with a standard deviation of 2.35. The utility here will calculate the average length of words for text that you copy and paste: <http://countwordsworth.com/lettersperword>

Although this is not always true, word length is used as a short-cut for the sophistication of texts (for example, children’s books have a shorter average word length than college Biology textbooks). Does the average word length that Presidents use differ from the English language as a whole? What assumptions did you make to answer this question? Do you think those assumptions are justified?

Note: To complete this assignment, you will have to take a random sample of the State of the Union speeches. Typically, a sample size of 30 is considered appropriate to apply the CLT. To get a random sample between a range of numbers, a – b, use the following code in excel: =RAND() \* (b-a) + a. *Please include the years for the speeches that your sample was based on somewhere in your answer.*

2. If we calculate 100 p-values below a .05 significance threshold, and reject the null hypothesis in favor of the alternative hypothesis, how many times do you expect we be wrong to reject the null? Please explain your answer, including stipulating what a p-value represents.