LING 609: HW 2
Due date: Tuesday, October 27th
October 20, 2015

1 Problem 1 (1 points)

Levy (forthcoming), exercise 2.9: For adult female native speakers of American English, the distribution of first-formant frequencies for the vowel [ɛ] is reasonably well modeled as a normal distribution with mean 608Hz and standard deviation 77.5Hz. What is the probability that the first-formant frequency of an utterance of [ɛ] for a randomly selected adult female native speaker of American English will be between 555Hz and 697Hz?

2 Problem 2 (5 points)

Load the data set lexdec from the languageR package. This data frame consists of data from a lexical decision experiment; consult the help page for lexdec to get more information about the contents of the data frame.

In this experiment, there are two groups: English speakers, and speakers of other languages. Your task in this problem is to determine whether the reaction times differ significantly by group. Note that the reaction times are logged in lexdec; you may either use the logged RTs or the raw RTs in your analysis, but you should justify your choice either way. In your analysis, you must:

i) Calculate each subject’s mean RT using either tapply or dplyy. With this information each subject’s mean RT, calculate the group means for native and non-native speakers, and the standard error of the mean for each group.

ii) Apply the appropriate t-test to test whether the means for the two groups are the same or not. In choosing and applying the appropriate t-test for this data, be sure to check your assumptions: are the data normally distributed? Are the variances of the two groups equal?
Present your a brief write-up of your analysis in the comments of your script. Do this in a clean, coherent fashion that would appropriate for a paper. In this write-up, present the results of your t-test, and state what conclusions, if any, you may make on the basis of that result.

3 Problem 3 (4 points)

The $\alpha$ level is conventionally set at 0.05 in linguistics and other social and behavioral sciences. The $\alpha$ level is type I error rate of your statistical test: the probability that you will reject the null hypothesis when the null hypothesis is in fact correct (i.e. the probability of a false positive). Using simulation, show that this is correct for the t-test. You may use either a one sample or an independent samples t-test in your simulation, though make sure that you choose a test that is appropriate for the data that you simulate. Demonstrate that this holds for a range of different values for $\alpha$.

EXTRA CREDIT (2 points): For extra credit on this problem, create a function to automate your simulations. Your function should take at least one argument, an $\alpha$ level, and return a single value, the type I error rate. To learn about how to make your own functions in R, consult this excellent tutorial at r-bloggers.com: http://www.r-bloggers.com/how-to-write-and-debug-an-r-function/.