

Homework 8

Due in class Tuesday, September 27th.

To read:

1. Read the excerpt from Nate Silver's *The Signal and the Noise*.

To hand in:

1. *For our Linguists*: Using the Corpus of Contemporary American English data on the website <http://www.wordfrequency.info/free.asp> and *assuming* that the probabilities of a word being in a sentence are independent (which is probably completely false) and *also assuming* that the probability that a word appears in a sentence is the same as the probability that it appears in the the whole corpus (also probably not true) find the probability that a sentence contains the two most frequent words in English 'the' and 'be'.
2. Building on 1. above, how would you calculate this probability if you didn't assume that the probabilities of 'the' appearing and of 'be' appearing were independent? Write both a verbal explanation and a formula where you have introduced symbols for the events and probabilities. [No need to calculate a number for this part because the website doesn't give you all the data you would need to do so.]
3. *For our Gamblers*: If you roll two 6-sided dice, what is the probability that you will get snake eyes (two 1's)? What is the probability that the sum of the two numbers on the face of the dice will be 7? [To get you started: List all possible rolls of the two dice in a table. Do this systematically so that you don't miss any possibilities.]