Statistics and Data Analysis in MATLAB
Kendrick Kay, kendrick.kay@wustl.edu

Homework 2 (covering Statistics Lectures 1 and 2)

To complete this assignment, prepare a MATLAB script called `homework2.m` along with any necessary accompanying function .m files. Then, run the MATLAB `publish` command (e.g. `publish('homework2.m')`) to run the script and generate HTML output showing the results. Turn in a print-out of the HTML output (e.g. from your web browser) and also a print-out of any function .m files that you write.

Hint: In your script file, place `%%` on a line by itself at each point where you want the HTML output to show figures and command-window text. Please note that your code should be commented (where necessary), including documentation of any functions that you write.

_____

Problem 0. Download the .mat file at http://artsci.wustl.edu/~kkay/psych5007/Homework2.mat (you will need this file to complete some of the problems below).

Problem 1. Draw a line plot that shows the Gaussian probability distribution with mean 50 and standard deviation 10. Label the axes.

Problem 2. The `data1` variable in Homework2.mat is a matrix of dimensions 3 x 20. Each row contains measurements obtained for a distinct group of subjects. We are interested in assessing whether there are any differences across the three groups. The metric that we will use is variability in the group means (i.e. compute the mean of each group and then compute the standard deviation across the three mean values). If this variability is high, that would suggest that the groups are different, whereas if the variability is low, that would suggest that the groups are the same. Use randomization to test the null hypothesis that all three groups come from the same underlying distribution. Run 10,000 simulations and calculate a one-tailed *p*-value. Echo the computed *p*-value to the screen.

Problem 3. Suppose we measure a group of 100 subjects before and after a certain manipulation. The `data2a` variable in Homework2.mat is a vector of dimensions 1 x 100 with the pre-manipulation measurement, and the `data2b` variable in Homework2.mat is a vector of dimensions 1 x 100 with the post-manipulation measurement. Quantify the manipulation effect by subtracting the pre-manipulation from the post-manipulation measurement. Then use bootstrapping to test the null hypothesis that the differences come from a probability distribution with mean 0. Run 10,000 simulations and calculate a two-tailed *p*-value. Echo the computed *p*-value to the screen.

Problem 4. Flip a coin *n* times and count the number of heads obtained. Repeat this 10,000 times and visualize the distribution of results using a histogram. Perform this simulation for different *n* values (*n* = 10, 100, 1000) and show the resulting histograms on three different subplots of a single figure. For each subplot, label its axes and provide a title.