Hypothesis Testing: Problem Set

1. Suppose you are performing a behavioral experiment where you measure the amount of time it takes for two groups of mice (wild-type and knockout) to find the hidden platform in a Morris Water Maze. Assume that you have 10 mice in each group and that after learning the escape latency is distributed as a gaussian.

   a. If the actual standard deviations are 10 seconds for both groups, the actual mean time for group 1 is 20 seconds and the actual mean time for group 2 is 30 seconds, what is the probability that you will detect the difference between the groups? The easiest way to do this is to simulate the data and see how often the differences come out significant.

   b. Suppose that the two populations are actually identical, and that both have a mean time of 25 seconds and a standard deviation of 10. Simulate 1000 random draws from times for the two groups and determine how often a two sample t-test returns a significance result at the default significance level. How is that number of incorrect outcomes related to the significance level?

   c. Suppose you were testing four pairs of identical groups with 10 mice each. What is the probability that one or more tests would result in a significant difference at the $p < 0.05$ level?

2. Run the matlab script **anova_test.m** (download from course website). Modify the data values so that the analysis produces

   a. A significant main effect of group but no other significant effects.

   b. A significant main effect of time but no other significant effects.

   c. Bonus: a significant interaction between group and time but no significant main effects.

3. Load the data from ps3_3.mat.

   a. What test should you use for these data?

   b. Are the data from the two groups different?

   c. Assume the data are from the experiment described in 1. What could the pattern of results imply?

4. Load the data from ps3_4.mat. Answer the same questions as in problem 3.

5. Load the data from ps3_5.mat. Answer the same questions as in problem 3.

6. Load the data from ps3_6.mat. There are two cell arrays, spikes1 and spikes2 representing the spike times of two neurons (in seconds) in a 50 trial experiment. spikes1{1} is the set of spike times for trial 1, spikes1{2} is the set of spike times for trial 2 and so on. A stimulus is presented at t = 500 ms.

   a. Compute the mean and variance of the firing rate in the baseline period (0 - 500 ms) for both cells.

b. Use the results from A to generate a histogram of z-scores. Use a bin size of 100 ms, and for each 100ms bin from t = 0 to t = 2 seconds, compute the z-score for the average rate in that bin.

c. Compare the raw data and the z-score data. Which cell is more strongly modulated by the stimulus? What interpretational issues does using Z-scores bring up?