**NUIN 408 – Homework #3 – Spring 2020**

Due: 4/29/2020 by 11:59 pm

**Problem 1**

You want to test a new drug to see if it is effective at preventing seizures in a rat model of epilepsy. You collect a small pilot data set from only 5 epileptic animals in control conditions (without the drug) so you can design the study. The variable you are measuring is seizure rate (per day). You monitor the 5 animals for a week and they give you the following values:

Seizures per day: 4, 5, 7, 2, 2

**a)** What are the mean and standard deviation of your pilot data? Fit the data with a normal distribution using the ‘fitdist’ function. Do you get the same values for the mean and standard deviation?

**b)** Now estimate the parameters of the normal distribution from which your data were sampled using a maximum likelihood estimator. Are the mean and standard deviation the same as the values you got in **a**? How do you explain any discrepancies? Use the parameter values in **b** for the rest of this problem.

**c)** You decide that your threshold for effectiveness is limiting seizures to twice per day in this model. You want to make sure that you have a 90% chance of rejecting the null hypothesis (that the drug was ineffective) if it indeed limits the animals to 2 seizures per day. Assuming you test the same set of animals before and after drug administration, how many animals do you need in the full study?

**d)** If you instead want to be 99% sure you do not miss the chance to reject the null hypothesis, how many animals do you need?

**e)** You only have the budget for 6 animals. If you still want 90% assurance that you will correctly reject the null hypothesis, how large an effect does the drug need to have to be detectable?

**f)** What is your statistical power with 6 animals for detecting an effect of the drug that limits seizures to 1 per day? What is the power for an effect that limits them to 3 per day?

**g)** Back at 90% statistical power, how many animals do you need to see a reduction to 2 seizures per day if, instead of the paired design, you have separate groups of control and drug animals?

**Problem 2**

Load SynapseData.mat. This file contains (fake) data from 100 voltage traces in a cell where glutamate was uncaged onto different synaptic locations. The values are in mV. The rows are each uncaging location and the columns are time points, from 0 to 50 ms in time bins of 0.1 ms.

**a)** Plot one of the first 10 traces with appropriate axis labels. Fit the trace with the equation below and report the values of you 3 parameters you obtain from the fit. Alpha is a scale parameter, d is a synaptic delay (in ms) and Tau is a decay time constant (in ms). You can use these as starting parameter guesses:

Alpha = 5; d = 10 ms; Tau = 15 ms.



**c)** Now fit this equation to each of the 100 traces and collect the parameter values. Also compute the R-squared value for each fit and collect those as well.

**d)** Do you notice any outliers among the fits? If so, explain how you found them, and remove those data points from your collection of fit parameters.

**e)** Plot histograms of the alpha, d, and tau parameters from your good quality fits. What can you say about the shapes of these distributions? Can you use chi-squared goodness-of-fit tests to test these conclusions?

**Problem 3**

Load FMRI\_data.mat. This is a (fake) data set representing BOLD signals from a simple experiment in which a subject was in the scanner for several minutes to establish a baseline signal in each voxel, and then the subject was shown an image of a face. The image was presented on 50 separate trials.

There are 2 variables in the mat file:

1) ‘voxels\_baseline’ is a vector of the mean baseline activity of each of the 1000 voxels, in arbitrary units.

2) ‘voxels\_faceImage’ is a matrix of size 50 rows and 1000 columns. This is the signal on each of the 1000 voxels for each of the 50 presentations of the face image.

The purpose of the data analysis is to look for voxels in which the presentation of the image had a statistically significant effect on the BOLD signal.

**a)**Use the “scatter” function in matlab to plot the activity of each voxel at baseline versus its mean activity during the image condition. What does this plot tell you? About how many voxels appear to respond to the image? Are all the responses increases in BOLD signal?

**b)**For the data from the first 3 voxels in the image data set, test the assumption that the 50 trials come from a Gaussian distribution. You can assume (for the sake of this problem) that the answer is the same for the rest of the voxels, but feel free to check.

**c)**Given your answers to parts **a** and **b**, choose a statistical test to decide, for each voxel, whether the face image altered the BOLD signal from its baseline level. Why did you choose this test?

**d)**Run the test for each voxel of the face data with the default value 𝛼 = 0.05. Report the number of voxels for which the null hypothesis was rejected. For each voxel, plot the p-value on a ‘semilogy’ plot. Plot a cut off-line at p = 0.05.

**e)**Does the number of ‘significant effect’ voxels you found in part **d)** match your intuition about the number of responsive voxels from part **a)**? If not, what could explain the discrepancy?

**f)**Repeat part **d** but with 𝛼 = 0.05/1000. What is now the number of ‘significant effect’ voxels? How does this number compare to your part **a** intuition?

**g)**Unlike in a real experiment, you now get to ask the oracle how many voxels actually responded to the image. In this case, you can trust the oracle (me), because I will post the code I used to generate the data after the homework is turned in. The oracle tells you the following:

14 voxels respond to the face image.

For 20 values of 𝛼 between 0.05 and 0.05/1000, plot (on the same graph), the false positive error rate and the false negative error rate for the face image. What conclusions can you draw about the appropriate choice of 𝛼 in this kind of experiment?