HW #4 (week 5)

- Due date: Wed, July 15 by 11:59pm EST. If you need extra time, just let me know.
- Download the data files to your local directory
- Send me both your script for analyzing the data (lastname_firstname_hw4.m) and the figures you make as .png files.
- Turn in the assignment by email at mlnguyen@princeton.edu

In this HW, you will be replicating a subset of analyses from my recent paper on teacher-student neural coupling during teaching and learning (link). In brief, a teacher was scanned in fMRI while teaching a lesson, which was recorded in-scanner. A group of 20 students were then scanned while watching the lesson. The students took a pre-test quiz to access their background knowledge prior to the lesson, and then took a post-test quiz after the lesson to measure learning.

Problem 1:

a. Load the "intact_behav.mat" file. This file two variables, `pretest` and `posttest`, that contain the scores for each subject. Scores range from 0 – 1. The order of data is by subject number (e.g. the score in `pretest(1)` and `posttest(1)` are for subject 1, the score in `pretest(2)` and `posttest(2)` are for subject 2, etc).

b. Using the appropriate statistical test, test for differences between pretest and posttest scores. State your null and alternative hypothesis and the type of test you are running.

c. Plot your results in an appropriate format. Remember to label axes, include error bars, and make the figure readable. Save your figure as a .png file (use the `saveas(gcf, 'filename', 'png')` function.

d. Describe your results from 1b and 1c.

Problem 2

The "roidata.mat" file contains the This file contains the fMRI response over time (a response timecourse), averaged across voxels within dorsal precuneous, a mid-line region of the brain (link, labeled in red as "PrC") that's involved in integrating information over time. Interestingly, research has also found that activity in the precuneous is "coupled" or correlated between speakers and listeners during communication: the listener's brain tends to mirror the speaker's brain. The strength of this neural coupling is measured using intersubject correlation, or ISC. ISC is calculating by correlating each listener's response time course with the speaker's response time course. The average of the resulting correlations is ISC. Here, you will extend these findings to neural coupling between a teacher (speaker) and students (listeners) during non-narrative teaching and learning.

a. Load “roidata.mat”. This file contains two variables: `student_data` is an nSub x time matrix containing the average precuneous timecourse for each student (each row is a different subject in the same order as the behavioral data) while they watched the lecture in the scanner. `teacher_data` is a 1 x time matrix containing the precuneous timecourse for the teacher while she taught the lecture in the scanner.

b. Calculate correlation between each student’s response time course and the teacher’s response time course. Keep track of these values for Problem 3. Calculate ISC by taking the mean of these correlations. Report both the mean ISC and the std.

c. On a single figure, plot the average student response time course and the teacher response time course in different colors. Include the ISC value on the figure, and don’t forget to label axes, add legends, etc. Save your figure as a .png file.

Problem 3:
Previous research suggests that the strength of coupling between speakers and listeners during story telling is related to how well they understand the story. Here, you will extend these findings to student-teacher coupling during learning.

a. Correlate the teacher-student correlations in precuneous from Problem 2 with the difference of pre-test scores from post-test scores from Problem 2. You should use each subject’s correlation with the teacher, not the ISC value. Report the r statistic and associated p-value.

b. Fit a linear regression model that predicts the improvement in learning score from teacher-student neural correlations.

c. Plot the results of this brain-behavior correlation in an appropriate format. Include the correlation from 3a, the “best fit line” from b, and the appropriate labels, etc. Save your figure to a .png file.

d. Describe your results from 3a-3c.

Bonus:
In practice, for Problem 3, we don’t use the p-value that MATLAB gives us. This p-value is calculated from the t-distribution, and in order for that to be valid, there are a number of assumptions that have to be met. Neural data often does not meet these assumptions, so using standard [parametric] statistics is sometimes inappropriate. Rather than assuming our r-value follows a t-distribution, we instead simulated a null distribution for our data using resampling.

In this case, our null hypothesis is that there is no relationship between student-teacher correlation and learning outcomes (r = 0). We want to calculate r-values under the assumption that the null is true. The way we do this is shuffling the order of values in one of our data vectors. Remember that your student-teacher correlations vector contains the values in order of subject number (s01, s02, s03, s04, …) and that your learning outcomes vector contains the values in the same order (s01, s02, s03, s04, …). By shuffling the order of one of those vectors, you're randomly comparing behavioral data from one subject to the neural data of another subject. We expect the correlation of these random comparisons to be close to 0.

We repeat this process many, many times (typically 1000-10,000 times), each time randomly shuffling one data vector and calculating the r-value. Most of the r-values we'd expect to be close to 0, but by chance sometimes they'll be bigger or smaller. Together, these r-values define a resampled null distribution. We can then use this null distribution to calculate our p-values: we compare the observed r-value to the null distribution, and count the number of shuffles that resulted in larger r-values. The p-value = #shuffles greater than observed value / total # shuffles.

For a bonus problem, implement this shuffling procedure and calculate a p-value for the correlation in 3c.