# Assignment 4 – Signal analysis, PCA and permutation tests

In this assignment, you will do a wide variety of things, on several datasets. Specifically, the script need to do the following things (you can assume that data and files are in the same folder as the code). Each spec below is worth 1 point, with partial credit for partial fulfillment.

1) Load the data file *soundSignals* into the workspace. Once loaded, the workspace will contain 3 variables representing 3 different sound signals that were sampled at CD quality for 2 seconds, as well as the specific sampling frequency, in *fs*. Make sure that the script plays each signal as a sound, separated by pauses (user input). Then determine the makeup (which components it consists of) of each signal by visualizing the signal in the time and/or the frequency domain and write in a comment what these signal components are.

2) Use the function *audioread* to recover the signal and sampling rate from the audio file "LOY.mp4". Play it in MATLAB and write in a comment what you hear. Visualize the signal with a spectrogram in a way that closely resembles the spectrogram in this article http://bit.ly/LOY2019S (the one that immediately follows the sentence "This is called a spectrogram."). Now play the same signal with a sampling rate that is 20% slower as well as 20% faster than the original sampling rate. Note in a comment what you hear.

3) Read the data file *LSRP.xls* into the workspace. This data represents scores on the Levenson Self-Report Psychopathy Scale, a commonly used screening test for Psychopathy, a personality disorder that manifests in terms of antisocial behavior and abnormal affect. The first 26 columns represent the 26 items on the test. Each item is rated on a 5 point scale from 1 to 5; rows represent participants. The last column represents "sex assigned at birth" (0 = female, 1 = male) of the participant in question. To get full score on this spec, calculate a total score for each participant and make a histogram of the empirical score distribution that also indicates where the empirical mean of this sample is. To do that, you need to take into account that some items (6, 14, 19, 22, 24, 25, 26) are "inversely coded" to avoid response bias, which means that a higher rating indicates **lower** degrees of psychopathic traits. For instance, on these items, a response of 5 on item 14 should count as 1, and a response of 2 on 24 should count as 4. So you need to recode these items first, to get an accurate total score. Finally, you need to find a way to handle missing data and make a plot of the raw data before and after you process data in this way.

4) Whereas the LSRP is a widely used instrument, its factor structure is still in dispute. Use the data from 3 and do a PCA. Make a scree plot to assess how many factors should be retained and note in a comment how many factors you think should be interpreted meaningfully, based on either the elbow or the Kaiser criterion. Finally, calculate and note (in a comment) how much of the total variance is explained by these factors.

5) The literature suggests that psychopathy is more common and more extreme in males, relative to females. However, much of this literature was recorded in institutionalized contexts (prison in particular), so there is the very real possibility of sampling bias. Investigate this question by using the community sample we already used in spec 3 and 4 (the participants are young, non-institutionalized US adults). Because psychopathic traits do definitely not distribute normally, classic significance tests might not be indicated. Instead, use resampling methods to perform a permutation test. Conceive of a reasonable test statistic and visualize how this test statistic distributes if there was no difference between these groups (the "null distribution") and calculate the exact p-value. Doing so, comment on what you determine in terms of potential gender differences in psychopathy.