

Lab 3

Program 1- (call this lab3.1.py)

This program will provide more experience with lists and loops. Read a spreadsheet (ss-dos.txt or ss-unix.txt) of comma-delimited data (i.e. commas separate the data entries in the same row). The spreadsheet has a large number of rows, and each row has an arbitrary number of points. For each row, calculate the mean and standard deviation. Write this to an output file that has a mean and standard deviation (comma delimited) on each line.

Instead of using the typical equation for standard deviation (which requires a mean), there is a convenient method for calculating the standard deviation across a list of N numbers, labeled x:

$$S = \text{sum} = \sum_{i=1}^N x_i$$

$$SS = \text{sum of squares} = \sum_{i=1}^N x_i^2$$

$$\sigma = \text{standard deviation} = \frac{\sqrt{N \times SS - S^2}}{N}$$

$$\bar{x} = \text{mean} = S/N$$

S and SS are calculated as running sums across the datapoints. After these are accumulated, the standard deviation and mean are calculated.

Program 2- (call this lab 3.2.py)

For lab 2, you translated an mRNA to its peptide. This week, do the same, but by using the modules from Biopython. Use the same GAPDH FASTA sequence file (GAPDH-dos.txt or GAPDH-unix.txt). You will need to install Biopython.