**MATLAB For Scientists**

**HW5**

**Instructions**

* Complete each question below in an **individual** script or function. Each individual script should be named according to: HW#Q#\_script.m, or HW#Q#\_fun.m where <#> indicates the corresponding HW number or question. These are the scripts that will be considered for grading.
* Submit all files (**including** the provided data, but **excluding** any tables/files that are saved as a result of the HW question) as a compressed .zip folder with the name **<HW#\_YourCodeName.zip>.**
* Each script **must** run entirely through without error. If you did not finish part of a question, comment it out so it runs.
* Suppress all intermediate outputs other than your answer to the question; only the answer should display on the command line.
* Absolutely no hard coding beyond the minimum specified.
* **All plotting should now conform to our guidelines of publication quality figures**
* **Everything must be commented. Uncommented codes get zero credit!**

**Problems**

**Question 1 (4 pts)**

Each HW5Q1\_microbiomePatient<#>.txt file contains the vaginal microbiota species composition dynamics over time of 4 different patients, from [this](https://www.ncbi.nlm.nih.gov/pubmed/22553250) paper. Every column in this table is a different bacterial species, and every row is a different time point. Samples were taken twice a week for 15 weeks total. Write a script that does the following:

* Uses a for loop to collect the *number* of species present in each patient that are non-zero at every time point
* Determines which time point has the greatest number of species, on average, across all patients
* Displays this time point, in weeks, in a complete sentence

**Question 2 (4 pts)**

Researchers in [this](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC534658/) study hypothesized that telomere length shortens over time in response to stress. To test this, they measured the telomere length in peripheral blood mononuclear cells isolated from healthy mothers of a chronically ill child at an increasing number of years following the initial diagnosis. HW5Q2\_TelomeresAndStress.csv contains years since diagnosis in one column and the telomere length in another. Write a script that does the following:

* Removes all data from any year that contained less than 3 measurements
* Plots years verses average telomere length with the remaining data. The plot should include a line connecting data points, and error bars corresponding to the standard deviation.

Hint: unique(x) gives you all unique elements in vector x.

**Question 3 (4 pts)**

The file HW5Q3\_ebola\_data.xlsx contains the number of confirmed, suspected, and probable Ebola cases over time for various countries. We’d like to determine if there was a significant drop in the mean number of all such cases in Guinea for the 5-month period of January-May 2015 compared to the 5-month period of August-December 2014. Write a script that does the following:

* States the null and alternative hypothesis as comments.
* Calculates the p-value using the correct t-test to test the question stated above.
* Displays whether the null hypothesis is accepted or rejected at a=0.05 in a complete sentence.

**Question 4 (4 pts)**

The file HW5Q4\_glioma\_labeled.csv contains the full glioblastoma data set that we reviewed in class. In total, there are 42 samples from 3 cell types: cancer glioblastomas (C-GBM), cancer oligodendrogliomas (C-GBM-O), and non-cancer glioblastomas (NC-GBM), which have 14 replicates each. We want to run a one-way ANOVA on individual genes of interest (which are arranged by column in the dataset), given the gene name. Write a function that does the following:

* Takes in (1) the full table and (2) a gene name as inputs
* Outputs all group pairs that show statistically significant differences using an alpha level of 0.05.

Then, write a script that:

* Calls this function using any representative gene
* Displays the significant pairings in a complete sentence.

Note that we should be able to change this to any gene in the table.

Hint: tableName.Properties.VariableNames gives you a cell array of the column names in a table.