**Put out the Fire! Exploring the Effects of Fire-Fighting Foams on Microorganisms using Statistics and Plots**

**Developed by T. Weathers
Chabot College, Fall 2020
tweathers@chabotcollege.edu**

**Objective:** Analyze a subset of data from a published research paper using MATLAB. This will focus on data manipulation in MATLAB, including reading in text files, for, if, and while statements, statistics, and plot generation.

**Background:** Consider the paper published in *Environmental Science and Technology* titled “Perfluoroalkyl acids inhibit reductive dechlorination of trichloroethene by repressing *Dehalococcoides*” found here: https://pubs.acs.org/doi/10.1021/acs.est.5b04854

In this paper, the authors found that perfluoroalkyl acids (PFAAs, the chemicals in things like Teflon and GoreTex) stop the growth of certain microorganisms. This allows other microorganisms to flourish.

**Data:** The data you are working with is directly from this paper. You are given the relative abundances (fraction of the total number of microorganisms observed) of different types of microorganisms. One set was exposed to PFAAs, another set was not. Each set consists of three different replicates, or repeat trials.

**Problem:** Check the statistical significance of each test, i.e., test whether there is a significant difference in relative abundance of each microorganism when it was exposed to PFAAs. We will use the Student’s T-Test, however, we must determine which type of T-Test to use. We will consider a final p-value of less than 0.05 to be significant by following the steps on the next pages. Be sure to read the instructions thoroughly before you begin coding!

**Procedure:**

1. The data is located in a separate text file. Input the data into MATLAB using the method of your choosing.
2. For the following calculations, it is helpful to complete them in a “*for*” or “*while*” loop, e.g. “*for i = 1:1:15*” where *i* will represent each row/microorganism
3. Calculate the means for each microorganism:
	1. For the case with PFAAs, find the average of PFAA\_1, PFAA\_2, and PFAA\_3 for each microorganism. Make this into an array called “meanPFAA”.
	2. For the case without PFAAs, find the average of No\_1, No\_2, and No\_3 for each microorganism. Make this into an array called “meanNo”.
4. Calculate the variance for each microorganism using MATLAB’s built-in variance function. Store the variance values in two arrays as described below:
	1. Find the variance between PFAA\_1, PFAA\_2, and PFAA\_3 for each microorganism. Save this as an array titled “PFAAVariance”
	2. Find the variance between No\_1, No\_2, and No\_3 for each microorganism. Save this as an array titled “NoVariance”
5. Create a column vector titled “FValue” for the ratio of (variancelarge)2/ (variancesmall)2, as in, the larger variance always has to be on the top. Hint: use an *if-*statement in MATLAB to determine which variance is larger and place it in the numerator.
6. Check if variances are considered equal or unequal:
	1. Using an F-test (details elsewhere) we know that our F-statistic is 19. If our calculated Fvalue from step 5 is smaller than 19, our variance is considered equal.
	2. Create a column vector that displays 1 for equal variance and 0 for unequal variance with each row corresponding to the microorganism.
7. Calculate t-test statistic, P
	1. We will be using the Statistics and Machine Learning Toolbox built-in function *ttest2* to calculate P
	2. Use *if*-statements to calculate the P value for each microorganism based on the following rules. Note that it will be helpful to store each P value in an array with rows corresponding to microorganism.
		1. If the variance is equal (Fvalue < 19) use the command:

*[H,P] = ttest2(x,y)*

where *H* indicates whether or not the hypothesis is null
 *P* is the P value
 *x* is the array of PFAA\_1, PFAA\_2, and PFAA\_3 for a microorganism
 *y* is the array of No\_1, No\_2, and No\_3 for a microorganism

* + 1. If the variance is unequal (Fvalue >= 19) uses the command:

*[H,P] = ttest2(x,y,’Vartype’,’unequal’)*

1. A P value of 0.05 or less is considered a significant difference. For example, if *Treponema* has a P value of 0.0385, there is a statistically significant difference between *Treponema* exposed to PFAAs and *Treponema* without PFAAs.
2. Have MATLAB print to the command window the names of the microorganisms that were significantly different between No PFAAs and With PFAAs. Hint: You can use the *find* command to find each row that is significant.
3. Create plots of the average relative abundances for each microorganism, with and without PFAAs
	1. Create a stacked bar chart with PFAA Presence or Absence on the x-axis and average relative abundance on the y-axis. This means that there will be two columns, and each microorganism will correspond to a specific color. Appropriately format the plot. Hint: You will need to put both mean values into a 15x2 array. To stack them appropriately, consider playing with transposes. See below for an example of what the final product should look like:



Figure 1: Example figure using a stacked bar plot

* 1. Create a different type of plot of your choice with the same data. Appropriately label the plot.

**Follow-Up Questions**

1. Consider the following sentence from the abstract of the paper referenced in this exercise:

Notably, there was significant repression of *Dehalococcoides* (8-fold decrease in abundance) coupled with a corresponding enhancement of methane-generating Archaea (a 9-fold increase) [upon exposure to PFAAs].

Note that “Archaea” refers to *Methanobacterium* in our data. Does your analysis support this conclusion?

1. We chose a p-value of 0.05 to establish significance. What happens if this threshold is higher? Smaller? Do you think a p-value of 0.05 truly represents significance?
2. Why do you think some microorganisms would be impacted by PFAA presence while others did not change?
3. Think about the two plots you generated. Which one does a better job of showing how the averages change upon exposure to PFAAs? Can you think of a better way to show this interplay?