Quantitative Methods Boot Camp Homework Day 2

1. Consider the following code:

```
x=[];
for i=1:2:10
   x(i)=i;
end
```

   a. How many times will the loop run?

   b. What will x look like when the loop is done?

2. In the following code, what's the chance that A will equal 1?

```
A=ceil(rand*100)
if mod(A,10)~=2
   A=1;
else
   A=2;
end
```

3. What is x going to look like?

```
x=[];
for i=1:10
   for j=1:10
      for k=1:10
         if i>j & j>k | (i==k & i==j)
            x(i,j,k)=1
         elseif mod(i+j+k,10)>9
            x(i,j,k)=2
         else
            x(i,j,k)=-1;
         end
      end
   end
end
```

4. Look up the help on the while loop. What does this program do? Think big picture.

```
a=round(rand*100)
p=0;
while (p==0)
   p=1;
   for x=2:a/2
      if mod(a,x)==0
         p=0;
      end
   end
   a=a+1;
end
a-1
```

5. In a single line of code, create a, a 10 by 10 array where all values are zero except along the diagonal where all values equal 5. (Hint: use the `eye` function (look this up with the MATLAB help)).

6. Starting with the a array you just created, and using a single line of code: to each element of the sixth row and eighth row of the array a, add double the value of the element in the sixth row, six column of array a.

7. Use the same array a from problems 5 and 6:

   a. Save a new array b1 as the array a to the power of the value in the fourth row, fifth column of array a. (x^y is x to the power y)

   b. Save a new array b2 as each element of the array a to the power of the value in the fourth row, fifth column of array a.

8. Using the array from the last 3 problems:

   a. What is the average value of each column of array a? (Try using the built-in MATLAB function)

   b. What is the average value of each row of array a?

   c. What is the average value of every third column (1st , 4th, 7th, etc.) of the array a?

   d. What is the average value of every other row (1st , 3rd, 5th, etc.) of the array a?

9. a. Make the following array

$$A=\begin{bmatrix}1 & 1 & 2\\ 2 & 7 & 7\\ 3 & 9 & 9\\ 4 & 7 & 3\end{bmatrix}$$

b. Now, make a new array B that is a normalization of array A (such that all numbers are between 0 and 1). This is done by first subtracting the smallest number and then dividing by the range (largest number - smallest number) of the array.

c. Use `hist` to view histograms of both A and B and verify that you did the right thing.

10. a. Write a function which takes 3 numbers (x,y,z). Have it create z random integers that are evenly likely to anywhere between x and y. Convince yourself that your function does what you want.

b. What is the effect of subtraction on the median, mean, and standard deviation?

c. What is the effect of division on the median, mean, and standard deviation?

d. What is the effect of removing the top 5 and bottom 5% outliers on the mean, median, and standard deviation of a normal distribution (`randn`)?

e. What is the effect of removing the top 5 and bottom 5% outlier on an exponential distribution (`exprnd`)?

11. Consider the following code:

```
A=[1 2 3; 4 5 6];
r=2;
s=A(r);
r=s;
clear
B=[7 8; 9 10];
t=B(2,1);
clc
u=B(3);
```

At the end of this, what is the value of the following variables?
A, B, r, s, t, u

12. What does the ' command do when placed after an array?

13. The file `marathon.xls` contains marathon training and performance data from a study on 47 runners (http://www.jhse.ua.es/jhse/article/view/190).

    a. The data is stored in a excel spreadsheet. The first column is average weekly training volume (in km per week) and the second column is marathon race pace (in seconds per kilometer). Make histograms of each column.

    b. For the group studied, what is the relationship between race pace and training volume (label your axes)?

14. File `bostonRain.mat` contains data about measured total rainfall (in inches) at various locations in Boston in the years 2003 to 2012.

    a. Load the data. Use `whos` to see what kind of array you are working with.

    b. You will notice that not all arrays are of the same length. This is because the later years include two more observation points that were not there in the earlier data. Write a script that compares the Station names of `rain12` and `rain3` and finds the stations that are new in `rain12`. (Hint: You might find the commands `strfind` and `cell2mat` helpful. You will also need to think about how to handle entries in a cell array).

    c. We only want to work with the weather stations for which observations exist for all of the ten years. Therefore, remove the extra data points from arrays `rain11` and `rain12`.

    d. Now, it would be handy to have all the data organized differently. One way to do this would be to have one array that just stores the data (with rows indicating the weather station and columns indicating years). Two different arrays could hold the information about the weather stations used and the years recorded, respectively. Implement this.

    e. By looping through the data array (or otherwise), find the first year when total rainfall (across all weather stations) exceeded 300 inches, and for that year, find the weather station with the maximum rainfall.