Mini Project #2 – Data Structure and Functions

Submitting This Mini Project

As this Mini Project requires you to create three functions, we need you to put them all in the same file when submitting them. Please follow these instructions for doing so:

1. Create a new script.
2. Copy and paste your answers for Part 1, Part 2, and then Part 3 into that script.
3. Save the script as ***yournetID***MP2.m, where yournetID is your Rutgers NetID, which has the general form aaa111. MATLAB will give you warnings in the editor. Do not worry about it.
4. Submit that file on Canvas.
5. Print it out to hand in during recitation.
6. Do not try to run the code in this .m file. It will not work.

Instructions

This Mini Project involves the creation of a structure that will hold a dataset and statistical information about the data set. First, let's describe the properties of our structure, `Data`. `Data` will have the following properties:

- `values`: an array of numbers in ascending order
- `mean`: the sample mean of all the numbers in value
- `median`: the sample median of all the numbers in value
- `variance`: the sample variance of all the numbers in value

Note that you can use any variable name for Data in the following functions, but the properties described above MUST be spelled and capitalized EXACTLY as above in order for you to get the last part of this project running. In addition to this, the functions we ask you to write in this project must have EXACTLY the names we ask for in order for the last part of this project to work. The last part is not for a grade, but it is fun for you.

Part 1: Data Structure Creation Function

Create a function named `createData`. `createData` should take one input:

- A number to start your data set off.

It should output a data structure, as defined above. `values`, `mean`, and `median` will all be your input number. `variance` will be 0.

Part 2: Adding New Numbers

Create a function called `addData`. `addData` should take two inputs in the following order:

- A data structure
- A new number to add to that data structure

It should output the data structure, with the new number included in the array contained by the `values` property. The number should be included in such a way that the array in `values` will always be sorted in ascending order. **You can assume that the array is already in ascending order when the data**

structure is given as an input. For example, if the array in `values` was `[1 3 5 7 9]`, and your input number was `6`, the new array would be `[1 3 5 6 7 9].`

Part 3: Updating Statistics

Note that you must accomplish everything in this part of the Mini Project without any built in functions that calculate means, medians, standard deviations, or sums of arrays. There are quite a few functions in MATLAB that do tasks like this. ***They are all forbidden here.*** *You should focus on using mathematical functions and operators that work on one or two individual numbers, loops, conditional statements, and array indexing.*

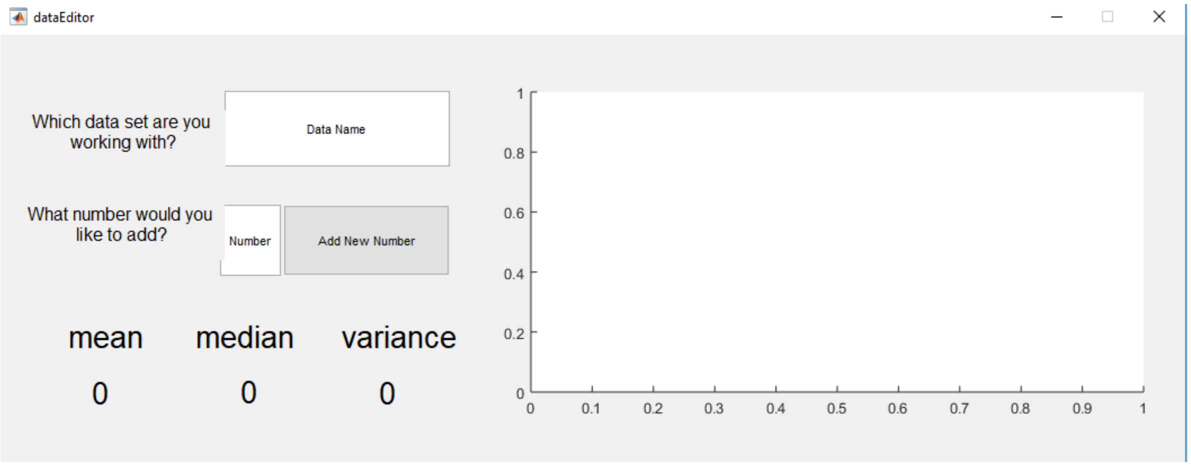
Create a function called updateStats. updateStats should take one input:

- A data structure

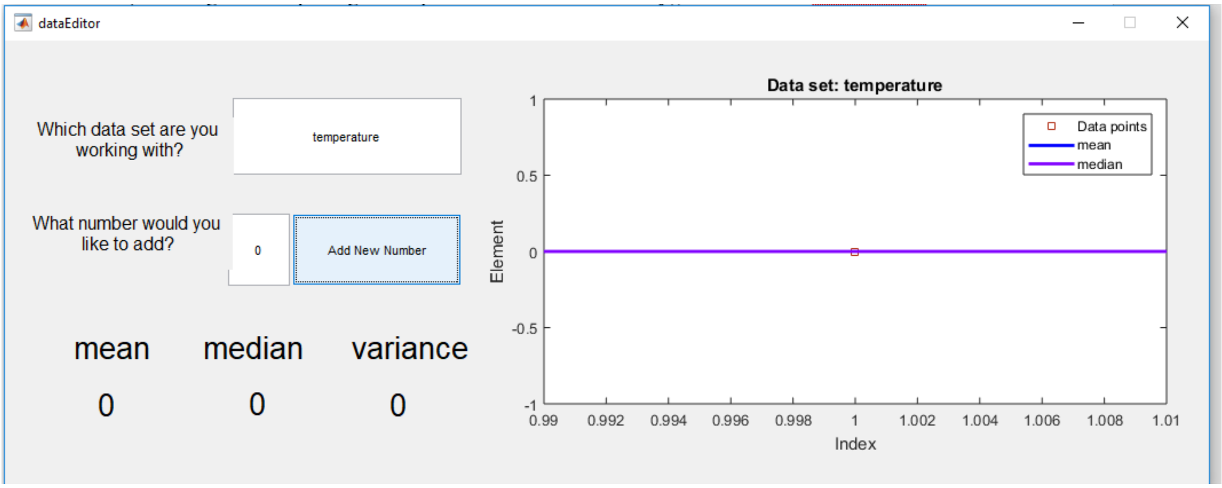
It should output a data structure with a newly calculated values for the `mean`, `median`, and `variance` properties. See **Appendix A** for information on how to calculate these values. **You can use the built-in MATLAB functions mean, median, and var to check your answers, but again, do not use them in the function.**

Part 4: Fun with a GUI

Once you are ***absolutely sure*** that the three functions you have written work as intended (including correct spelling and capitalization where necessary), download the dataEditor.m and dataEditor.fig files from the Mini Project link into your default MATLAB folder. Then, in the command window, type `dataEditor`. It will open a graphical user interface (GUI) that looks like this:



This GUI will call your functions to create and edit a data set, then display the statistics of that data set along with a plot. Try typing the data set name ***temperature*** in the first text entry box, and the number ***0*** in the second text entry box. Then press the “Add New Number” button. You will see this:



dataEditor

Which data set are you working with? temperature

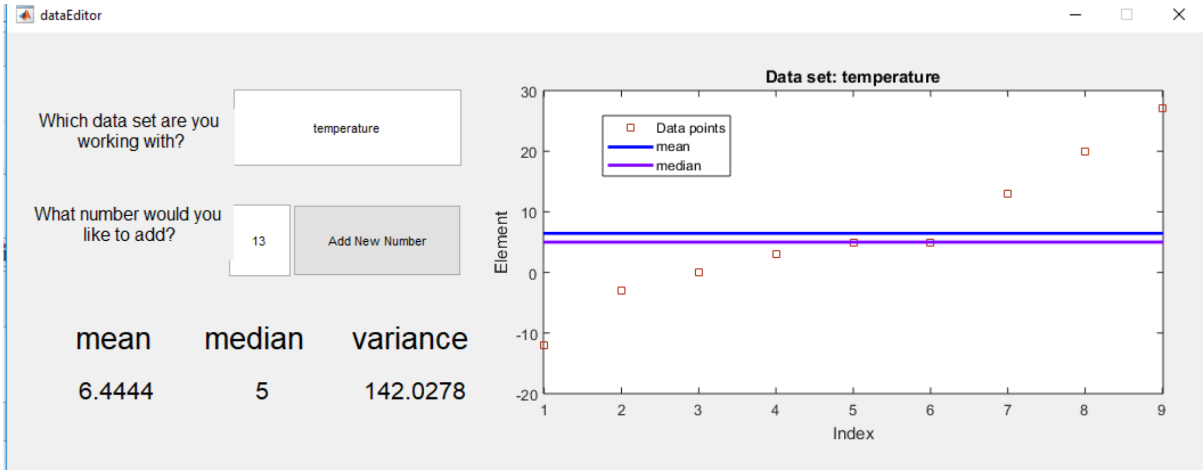
What number would you like to add? 0

mean 0

median 0

variance 0

If you continue to add new numbers to the data set, it will keep updating the statistics and the plot. In the next picture, I have added the numbers -3, -12, 20, 3, 5, 5, 27, and 13:



dataEditor

Which data set are you working with? temperature

What number would you like to add? 13

mean 6.4444

median 5

variance 142.0278

You can create different data sets, and keep adding new numbers. The data is saved to your computer, and will not be lost with the addition of new numbers. Play around and have fun! **This is an example of a data analysis tool you could build yourself in MATLAB with the concepts you will learn throughout the rest of this class.**

Appendix A: Calculating Array Statistics

Mean

For a given array, we calculate the mean, μ, with the following formula:

$$\mu = \sum_{i=1}^{N} \frac{x_i}{N}$$

That is, the sum of all elements, x_i, in an array of length N, divided by N.

Median

The median is defined as the middle value of an array. It should have exactly the same number of values that are higher and lower than it within the array. It is calculated differently for arrays of odd and even length. For example, for the odd lengthed array:

[0 1 1 2 3 5 8 9 9]

the median is a value from within the array:

[0 1 1 2 **3** 5 8 9 9] → 3

Which has 4 values less than it and 4 values greater than it. For the even lengthed array:

[2 3 4 10 11 12]

the median is the average of the two values in the middle of the array:

[2 3 **4** **10** 11 12] → (4+10)/2 = 7

There are 2 numbers less than 4, and 2 numbers greater than 10, so these are the middle numbers of the array that we use to calculate the median. Note that the numbers before and after the median can be the same as the median. The median of:

[1 2 2 2 2 2 2 2 2 2 3]

is

[1 2 2 2 2 **2** 2 2 2 2 3] → 2

Variance

For a given array, we calculate the variance, σ^2 with the following formula:

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N - 1}$$

That is, the sum of the squared differences between all elements x_i and the mean μ divided by the number of elements in the array minus 1. Note that the variance of an array with a single number reduces to 0.