CS 122 – Lab #3 – Simple statistics

Part 1: Mean and standard deviation

In statistics, we often have a list of data values that we want to be able to concisely describe. For example, how could we describe the list of numbers (1, 2, 8, 9) ? Besides reporting the size of the list, we are mainly interested in two things: some measure of "central tendency", such as a mean (i.e. average), and some measure of how dispersed the values are. The most common way to measure the dispersion of values is the standard deviation.

Consider these two lists: (1, 2, 8, 9) and (5, 5, 5, 5). Both lists have the same mean. But we want to be able to say that the first list has its values more spread out than the second list. This is what the standard deviation measures.

Calculating the mean is easy: add up all the values in the list and then divide by how many there are. But the formula for standard deviation goes like this:

$$standard\ deviation = \sqrt{\frac{\sum_{i=1}^{n}(x_i - mean)^2}{n}}$$

Here is the step-by-step procedure for calculating the standard deviation, along with a worked example for the list (1, 2, 8, 9). Remember to use real-numbers (not integers) throughout your calcualtions.

1. First, calculate the mean. In our example, it is (1 + 2 + 8 + 9) / 4 = 5.0
2. For each value in the list, calculate its deviation from the mean. In other words, subtract the mean from each value. In our example, we would obtain these deviations: –4, –3, 3, 4.
3. Square each of the deviations. In our example, we have 16, 9, 9, 16.
4. Add up all the squared deviations. In our example, this is 16 + 9 + 9 + 16 = 50.0
5. Divide by the size of the list. In our example, this is 50 / 4 = 12.5
6. Take the square root to get the final answer. $\sqrt{12.5}$ is approximately 3.54, and this is the standard deviation of our example.

Now you are ready! Write a program called Stat.java that will calculate the mean as well as the standard deviation of a list of values entered by the user. The user will supply the values at the command line, one at a time. The size of your array will be the same as the number of command-line arguments. You will need to convert each command-line argument into a `double`. Calculate and print out the mean and standard deviation.

Example run:

java Stat 1 2 8 9

The mean is 5.0
The standard deviation is 3.5355339059327378

Part 2: Median

Let's also tell the user what the median of the array is. This is accomplished in two steps:

1. Sort the array. Fortunately, Java has a built-in method for sorting arrays. You guessed it: It's called sort(). Here is how you use it. First, at the top of your program, include this statement:

   import java.util.Arrays;

   Inside the body of your program, when you are ready to sort the array, the command for sorting the array a in ascending order is:
   Arrays.sort(a);

2. Next, you need to find the "middle number" in the sorted array. If the size of the array is odd, then there is no confusion. But if the array size is even, we define the median as the average of the two numbers in the middle.

   For example, suppose the array has 1, 2, 3, 4, 8, 15. Then the median is the average of 3 and 4. As another example, suppose the array has just 1, 2, 3, 4, 8. Then the median is 3.

   Print the median as a 3rd sentence at the end of your output.

When you are done, please have your work checked by the instructor or lab aide.