# **Summary Statistics**

### Problem ID: a05p04summarystatistics

The average and standard deviation are incredibly useful summary statistics. The average, which is also known as the mean, tells us what the "typical" value is for a set of data whereas the standard deviation tells us how close or far away from the average the other values tend to be.

The average  $\mu$  of n numbers,  $x_1, x_2, \ldots, x_n$ , is easy enough to calculate:

1. Sum up all the numbers,

$$x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$$

2. Divide the sum by n,

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

Once the average has been calculated, the standard deviation  $\sigma$  is obtained by:

- 1. Squaring the difference between each value  $x_i$  and the average  $\mu$ , that is  $(x_i \mu)^2$ .
- 2. Summing up all those differences,

$$\sum_{i=1}^{n} (x_i - \mu)^2$$

3. Dividing the sum by n,

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

4. Taking the square root,

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

Notice how these steps resemble an algorithm ... but this is too easy.

To make this more interesting, let us implement an algorithm that computes the *running* standard deviation. A running standard deviation is calculated for every data point as it arrives. This can be useful for very large data sets or when values are arriving in real-time (e.g. sensor measurements).

The program should repeatedly ask the user to type in an integer value and then print out the cumulative moving average and the standard deviation for each value that the user enters. When printed, the summary statistics should be rounded to the 2nd decimal.

The WikiPedia article on standard deviation lays out how these calculations can be carried out, albeit expressed using mathematical notation.

We propose that you tackle this problem one step at a time:

- 1. Read the WikiPedia article, think about how step-wise calculations that compute the value at step k by using values from step k-1 could be implemented using a loop. (If you are stranded, don't be afraid to ask a teacher or a fellow student for help.)
- 2. Start by tackling the simpler problem: calculating the average. If you can solve that, then the rest becomes much easier.
- 3. Extend your solution to also update the standard deviation in every iteration of the loop.

NOTE: The program should calculate the population standard deviation, not the sample standard deviation.

#### Input

Input starts with one line containing one integer n, where  $0 \le n \le 10\,000$ . This input is repeated until n is -1.

## Output

Output consists of two lines for each number n that was input. The first line containing a float m, the average of all numbers input so far, and the second line containing a float s, the standard deviation of all numbers input so far.

Sample Input 1	Sample Output 1
Sample mput i	Sample Output i

•	•
2	2.00
4	0.00
4	3.00
6	1.00
4	3.33
-1	0.94
	4.00
	1.41
	4.00
	1.26

### Sample Input 2 Sample Output 2

2	2.00
4	0.00
4	3.00
4	1.00
5	3.33
5	0.94
7	3.50
9	0.87
-1	3.80
	0.98
	4.00
	1.00
	4.43
	1.40
	5.00
	2.00