TWO SIGMA

Research Technical Skills Test

Welcome to the Two Sigma Research Technical Skills Test.

Please read the following carefully as your solution may be rejected if you do not follow these instructions.

1. This is a timed test. We expect you to take no more than **3 hours** to complete this test from the time you receive this message to the time you email us your solution. Email us your final solution when you are done. There are no bonus points for turning your solution in early, so please check your work thoroughly if you have time remaining. If you have problems with your development environment, please notify us immediately by calling 646-292-6049.

2. We will accept solutions in Java (JDK 6 or earlier), C++ (ANSI C++ 14882-2003), C (ANSI C 99), Perl (5.10.0 or earlier), and Python (2.6.4 or earlier). You may only use standard, operating system neutral libraries and you must write your code to be portable. You do not need to use the same language for all problems.

3. For Problem 1, your solution will be judged primarily for its correctness, with attention to code clarity, design and efficiency. This problem requires that your program reads input from Standard Input (stdin) and prints output to Standard Output (stdout). Please leverage your favorite search engine if you are unfamiliar with standard I/O streams. **Correctly processing the input and output to your code is an essential part of this problem.**

4. For Problem 2, we are interested in the creativity and rigor of your analysis in addition to the correctness of your methodology and code.

5. If you have questions about the problem set, please use your best judgment and carefully document your assumptions in your code.

6. Please reply to this email when you have your final solution to the two problems. Create a separate .zip, .tar or .tar.gz archive for each solution and attach them to the email before sending. Do not include any binaries, only source code, as this might trigger our spam filters.

Thank you for participating in the technical skills test!

TwoSigmaResearchCodeTest $Revision: 1.4$

Problem Statement #1: Sequence Arithmetic

Assume an array containing a sequence defined such as:

$$\boldsymbol{a_0 = x}$$
$$\boldsymbol{a_i = Ka_{i-1} + L}$$

Write a program to

- Find K and L given such a sequence provided to you.
- Determine whether a sequence provided to you has such a relationship between $\boldsymbol{a_m}$ and $\boldsymbol{a_{m+1}}$. You may assume that $\boldsymbol{a_0}$ is always an integer, and that only addition / subtraction of L and multiplication / divisions of K are allowed.

Input:
The first line of input will contain 1 or 2, representing the operation desired: 1 for “find K and L” and 2 for “determine whether the relationship is of the right form”.

Following that will be a space-separated list of numbers. The first will be an integer ($\boldsymbol{a_0}$) but the others may not be. Numbers can be positive or negative. The line will be at most 1000 characters long.

Output:
If 1 is specified in the first line of input, then the program must output K and L to 2 digits of precision and separated by a space.

If 2 is specified in the first line of input, then the program must output either “true” or “false” – “true” if the given sequence does follow the relationship specified.

Constraints:
The code you submit must take input from stdin and produce output to stdout as specified above. No other output is permitted. You can assume the input will be valid.

Problem Statement #2: Data Analysis

Write a program to download as many NFL data files as you choose from http://www.repole.com/sun4cast/data.html. Attached to this e-mail we have included the data, as well as the script we used to download it, as an example and in case you want to use it for additional downloads.

Put the files in a directory structure of your choosing and make a zipfile or tarball. Send this back along with your code test.

After downloading the files, use the data to test a hypothesis of your choice, and write an analysis of your result in a file called README. For example, some football analysts claim home teams have an advantage over visiting teams. Others claim teams that are doing poorly will start doing better (and vice versa). Can you show either of these is true?

Please think about the following as you form your response:

- What is the hypothesis you are testing, and why does the metric (or metrics) you have chosen make sense to test the hypothesis?
- What are your assumptions going into the analysis?
- How much data are you working with? Is your result significant given the dataset?
- If you had more time, what extensions might you want to explore?

Please include the **all** of the code you wrote for this problem, particularly code used to retrieve and analyze the data.